RICE UNIVERSITY

Evolutionary fates within a microbial population highlight an essential role for protein folding during natural selection

by

Matthew Isaac Peña

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Yousif Shamoo, Thesis Advisor
Associate Professor of Biochemistry & Cell Biology

Jonathan Joff Silberg, Committee Chair
Assistant Professor of Biochemistry & Cell Biology

Kathleen M. Beckingham
Professor of Biochemistry & Cell Biology

Ronald J. Parry
Professor of Chemistry and Biochemistry & Cell Biology

Jeffrey J. Tabor
Assistant Professor of Bioengineering and Biochemistry & Cell Biology

HOUSTON, TEXAS December 2011

Abstract

Evolutionary fates within a microbial population highlight an essential role for protein folding during natural selection

by

Matthew Isaac Peña

The fitness function developed in this thesis directly links the physicochemical properties of an enzyme to evolutionary fates in a quantitative and predictive manner through a comparative study of empirical and simulated data. The success or failure of organisms during evolution is dictated by changes in molecular structure that give rise to changes in fitness revealed by evolutionary dynamics within a population. While the conceptual link between genotype, phenotype and fitness is clear, the ability to relate these in a quantitative manner remains difficult. I show here that predicting success during adaptation can depend critically upon enzyme kinetic and folding models.

We used a 'weak link' method to favor mutations to an essential, but maladapted adenylate kinase gene within a microbial population that resulted in the identification of five mutants that arose nearly simultaneously and competed for success. The unique catalytic role of adenylate kinase *in vivo* is to maintain adenylate homeostasis by catalyzing the reaction: ATP + AMP ↔ ADP. The stabilizing substitutions retained this essential function and were shown to be necessary for viability at higher temperatures. Physicochemical characterization of these mutants demonstrated that, although steady-state enzyme activity is important, success within the population is critically dependent

on resistance to denaturation and aggregation thus emphasizing the importance of proper folding in adaptation.

In vitro activity is a product of critical catalytic and folding pathways, and hence is a valuable proxy for fitness. A fitness function relating *in vitro* measurements of enzyme activity and reversible and irreversible unfolding to growth rate must impose an activity threshold above which there is no added fitness benefit in order to reproduce *in vivo* evolutionary fates in an *in silico* population. The fitness function thereby links organismal adaptation to the properties of a single gene. Understanding the physical basis for adaptation of an organism is the first step in the development of approaches that can accurately model, and someday predict, the manner in which organisms would respond to new antibiotics and improve upon the current clinical regimens.

Acknowledgments

The work presented herein would not have been possible or as rewarding without the support of many people over the course of my graduate studies. For my family and friends, the faculty and funding, I am thankful.

I thank my parents, Manuel and Anna, for always nurturing my curiosity and supporting my pursuits. I'd find refuge away from the laboratory with home cooking back in Austin, always bringing back as much food as I could upon my return. I would also like to thank Amy Davenport for her unyielding love and companionship despite the large distances that separated us over these five years.

I want to thank many of the faculty, in particular Kevin MacKenzie, Joff Silberg, Matthew Bennett and John Olson, for always having an open door and entertaining my questions. Kevin challenged me to think critically and deeply about the problem at hand, typically by leaving me puzzled during lab meetings. Whenever I'm in seminar or lab meeting, I always try to channel my inner Kevin. Joff always presented an impeccable work ethic and a state of constant productivity at the cutting edge. He once told me during my rotation, "PCR works while you work." I've shared this nugget of wisdom with many of my peers and undergraduates over the years. Matt Bennett was immensely helpful with shoring up the mathematics behind modeling a microbial population and the adenylate homeostasis pathway. Matt provided much support in making the transition from graduate student to postdoctoral research associate. Dr. Olson is as compassionate as he is inquisitive. He is as much of a role model for students becoming scientists as for them becoming adults. Dr. Olson provided critical assistance in developing the unfolding pathway necessary to fit the enzyme activity data at high temperatures.

I would like to thank my thesis committee, Joff Silberg, Kate Beckingham, Ron Parry and Jeff Tabor. Each deserves praise for being supportive of my efforts and providing constructive criticisms when necessary - both vital to my growth as a scientist. I am especially grateful to Kate for her assistance in proofing this thesis.

I have had the rare good fortune of being surrounded by coworkers, many of which I will not name here, I also call good friends. They have made graduate school a truly enjoyable experience. My treasured memories will be of the time I spent with them, and not the many hours I worked over a water bath with pipette in hand. Aaron Collier, Corwin Miller and Andres Benitez, I will miss our late night food runs, summer time snowball fights, and gaming late into the evening. You provided some semblance of normalcy when everyone else had gone home to sleep. Kasia Walkiewicz, we joined the lab together, and I was fortunate to find a true friend in the person sitting across the lab bench from me who I could confide in and commiserate with over these past years. I'd be remiss not to thank many others, especially Milya Davlieva, Tom Guu, George Zhao Huang, Chun-Jen Lin, Erol Bakkalbasi, Ian White, Sarah Wu, Chacko George, Chad Hruska and Jakub Otwinowski, for their friendship and assistance over the years.

And finally, Dr. Shamoo, it has truly been a pleasure to work with you. I always looked forward to our discussions about science and other trivial matters. You demonstrated improved patience over the years, and I'd like to think I contributed to that and a few grey hairs with my quest for smaller error bars and my speed, or lack thereof, in writing. You have cultured a positive laboratory environment, and that reflects your unique enthusiasm and guidance. I can only hope that some of this has rubbed off on me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

adk	Adenylate kinase, gene
AK	Adenylate kinase, protein
Ap_5A	P^1, P^5-di(adenosine-5') pentaphosphate
CD	Circular Dichroism
C_i^J	Flux control coefficient
CV	Column Volume
DSC	Differential Scanning Calorimetry
EC	Adenylate Energy Charge
EC number	Enzyme Commission number
FPLC	Fast Protein Liquid Chromatography
H	Enthalpy
K'_{eq}	Equilibrium constant
k_{cat}	Catalytic constant
K_M	Michaelis constant
$K_{M(obs)}$	Observed Michaelis constant
K_{unf}	Unfolding equilibrium constant
MWCO	Molecular Weight Cutoff
OD	Optical Density
PDB ID	Protein Database Identification code
S	Entropy
T	Temperature
T_m	Melting temperature
T_0	Melting temperature in absence of ligand
v	Reaction velocity
Y_U	Fraction of unfolded protein
ΔC_p	Change in heat capacity
$\Delta G'^\circ$	Standard Gibbs free energy
$\Delta G^\ddagger$	Activation energy
ΔG_p	Phosphorylation potential
ΔG_{unf}	Gibbs free energy of unfolding

Chapter 1 – Introduction

The question of the adaptive significance of enzyme polymorphism can be restated: to what degree is it possible to describe a molecular mechanism for fitness differences among enzyme variants?
Koehn, Zera & Hall, J. G. (1983)

In Nature, evolution occurs through the continuous adaptation of a population to its environment. Changes in protein structure at the atomic level are the basis for protein evolution, and in many cases these changes are reflected in altered phenotypes and fitness. A more precise understanding of the biophysical origins of fitness is central to one day predicting evolutionary outcomes. With the work described hereinafter, I will answer the question proposed by Koehn, Zera & Hall (Nei & Koehn, 1983) by demonstrating that enzyme activity can be interpreted as a proxy for fitness and that this relation can then be used to model the evolution of enzyme variants within a microbial population.

Continuous evolution of bacterial populations can be used to determine mutational pathways and adaptive mechanisms of increased fitness under natural selection. The complexity of adaptation can be reduced and the evolution of single genes studied by using a 'weak link' approach.[1] Previously, a homologous recombination strategy was used to replace the chromosomal copy of the essential adenylate kinase gene (*adk*) of the thermophilic bacterium *Geobacillus stearothermophilus* with that of the mesophile *Bacillus subtilis* (Couñago & Shamoo, 2005). Recombinant

[1] I suggest the weak link reduces the complexity of adaptation because improvements in the single gene product provide the greatest returns in fitness. This limits the complexity arising from epistasis between adaptations in multiple interacting genes. Changes in the one gene can then be attributed to differences in fitness therefore strongly coupling genotype (and performance of enzyme variants) to phenotype (in this case fitness).

G. stearothermophilus cells that expressed only *B. subtilis* adenylate kinase (AK_{BSUB}) were unable to grow at temperatures higher than 55 °C due to heat inactivation of the mesophilic enzyme and consequent disruption of adenylate homeostasis (Couñago & Shamoo, 2005). Continuously growing populations of bacteria were then subjected to selection at increasing temperatures (from 55 to 70 °C) that favor changes in the one gene not adapted for thermostability, *adk*. During the course of selection, the population was sampled and intermediates of adaptation were observed as mutations to *adk*. Mutations that improved AK function, tightly linked to organismal fitness, were favored by natural selection. Changes to AK activity and thermal stability resulting from mutations had direct consequences for cellular fitness. These enzyme activities and stabilities were examined to determine how the mutant populations traversed the adaptive landscape to increased fitness (Couñago *et al.*, 2006). The weak link approach tightly couples adaptive changes within the genome to changes in fitness. The biochemistry of adenylate homeostasis and results of weak link evolution are discussed in depth in Chapter 3.

The main objective of this thesis is to determine how the activity of adenylate kinase variants corresponds to fitness of the strains carrying the respective alleles within a polymorphic microbial population. This requires accurate descriptions of enzyme activity and stability for enzyme variants between 55 and 70 °C. The methods for determining these characteristics of enzymes, activity and stability assays, are described in Chapters 4 and 5.

In Chapter 6, I will show that reversible and irreversible folding pathways must be accounted for to accurately determine *in vitro* kinetic parameters (K_M and k_{cat}) of AK at temperatures where a significant fraction of free enzyme is unfolded. In addition to this, I

show that a nonlinear relation between activity and fitness is required to reproduce the *in vivo* evolutionary fates of the population within a simulation.

I have modeled fitness as a function of *in vitro* enzyme activity, which is partially a function of enzyme stability, and the application of a threshold that provides an upper limit on fitness. I hypothesize that an activity threshold exists above which no added fitness benefit is attained (the 'physiological threshold'). However, as activity falls below this threshold AK becomes rate limiting and fitness is negatively affected. This model can successfully reproduce frequencies of mutants in a polymorphic population, including the transient success of three minor mutants and order of disappearance from the population, given only *in vitro* data and allowing for the activity threshold to be fit to the observed outcomes. An appealing aspect of our fitness function is that it permits an evaluation of specific and quantitative aspects of protein stability and activity relative to evolutionary fates.

I further develop the fitness function in Chapter 7 by replacing the strict physiological threshold with a Hill-like relation to model fitness as a function of enzyme activity. I then show that this improved function has qualities that match the response of the adenylate homeostasis network to changes in AK activity. The culmination of this work is not the first time enzyme activity has been linked to fitness (see Chapter 2). However, the work herein represents the first use of fitness descriptions for enzyme variants used to model the success of each during an evolutionary process.

In vivo, diversity within a population is generated by a variety of mechanisms that span single nucleotide changes to genome-wide rearrangements and horizontal gene

transfer. However changes are generated within an organism, it is the physicochemical characteristics of the resulting macromolecules and their resultant changes in the fitness of the organism that are the 'grist for the mill' of natural selection. While our system links the physicochemical properties of adaptive changes that increase stability, the principles apply equally to those changes that might decrease stability of the ensemble either through mutation or translational errors. Thus, regardless of how protein diversity is generated, evolutionary dynamics will likely be strongly coupled to stability and function.

Chapter 2 – Physiological basis of fitness

If this (Fisher) hypothesis is untenable what alternative is there? I believe that little progress can be made without development of a physiological theory...whatever the evolutionary mechanism by which a particular gene complex has been reached, the state of dominance of all genes in the complex must always have a completely physiological explanation.
Sewall Wright (1934)

The object of this thesis is to demonstrate that evolutionary dynamics within a microbial population can be reproduced based on the *in vitro* enzyme activity and folding characteristics of adaptive mutants during selection. The historical thread described herein joins researchers responsible for the critical leaps in thought that have contributed to a mathematical framework for describing the physiological basis of fitness and those who have developed methods of studying adaptation within microorganisms. Undoubtedly the contributions of many other are important, though I have tried to be as inclusive as possible.

2.1 From Darwin to the modern synthesis (1859 – 1942)

Charles Darwin and Alfred Russel Wallace independently arrived at the same conclusion: all species, past and present, are related by common ancestry. Their shared hypothesis is that all present species – meaning all distinct types of organisms – originated from preexisting species, populations are not static, but may and do change over generations (Darwin & Wallace, 1858). Darwin called this phenomenon "descent with modification," which would come to be known as the theory of evolution (1859). The strength of Darwin and Wallace's theory resided with the proposed process of natural selection, which could explain the dynamism of populations and the incredible

diversity of all living things. Evolution by natural selection occurs whenever three conditions are met: phenotypic variation exists among the individuals of a population, the variable traits are heritable, and these heritable traits affect the average reproductive success of the individual – its fitness. Natural selection enriches the population with the genotypes of those individuals with increased fitness. Though powerful, the theory could not provide the source of variation on which selection acted.

The work of Mendel, a contemporary of Darwin with whose work he was familiar (De Beer, 1964), provided a theoretical mechanism explaining the heritability of variation of the form of dominant and recessive traits or "characters" in the formation of hybrids (1866). The traits that are transmitted in the course of hybridization and constitute the characteristic appearance of the hybrid were said to be 'dominant' and not intermediate in appearance, while those traits that contributed little were called 'recessive'. Providing a theoretical framework to describe the consistent reappearance of the recessive trait to a set proportion of the progeny of hybrids represented one of Mendel's unique contributions to the study of genetics. However, Mendel's work was largely forgotten, and it was not until its rediscovery at the turn of the century (Correns, 1900; De Vries, 1900; Tschermak, 1900) that the laws of inheritance were verified with studies in peas (*Pisum sativum*) as well as other plant varieties. The physical material of heredity, the chromosome, would soon be proposed by Theodor Boveri (1902) and Walter Sutton (1902), as the chromosomes were seen to undergo segregation and independent assortment according to the Mendelian rules of inheritance.

Ronald A. Fisher would later bridge natural selection and Mendelian inheritance, showing in works culminating with *The Genetical Theory of Natural Selection* that these

concepts were not antagonistic but in fact compatible (1930). Specifically, Fisher illustrated how Mendelian inheritance can produce heritable, gradual variation on which natural selection can act. The early 'presence vs. absence' theory of dominance proposed by Bateson (1909) suggested recessive mutations result in loss of a gene or gene product conferring the positive trait observed in the dominant form. However, this theory failed to explain why most mutations are recessive relative to the wild type (and the lack of intermediate phenotypes for heterozygotes), how dominant effects could arise from known losses of genes, how mutations revert, and how evolution is perpetuated if variation by mutation occurs only through loss (a shortcoming that Bateson acknowledged). Fisher, in his effort to explain the dominance of wild type over the majority of recessive mutations, suggested that the influence of a mutation on the organism's external characteristics were modified by additional hereditary factors, which were also under pressure of natural selection (Fisher, 1928). A shortcoming of this early explanation of dominance was that it was inconsistent with the view that evolution occurred by selection of a series of mutations which *improve* viability relative to the wild type progenitors (and not all mutants are recessive, only the majority).[2] John B. S. Haldane and Sewall Wright further weakened Fisher's modifier theory by showing mathematically that the selection rate for such modifiers is too slow and thus improbable to explain the evolution of dominance in all but rare cases (Wright, 1929; Haldane, 1930). Within these critiques lay the seeds of a physiological theory of dominance (predating metabolic control analysis by nearly 40 years), where the focus, now wrested from modifiers, would be placed on genes and their successive products. To Wright, it

[2] Fisher erred because he assumed, falsely, that heterozygotes should produce an intermediate phenotype, which Wright showed was not necessarily the case.

was apparent that phenotype was related to the quantity of product made by catalyzed reactions (Figure 2.1).

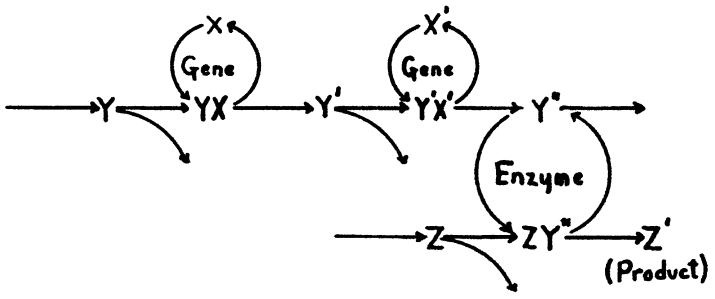


Figure 2.1
Wright's interpretation of successive gene products includes the biomolecular interaction between substrate (Z) and enzyme (Y") to form product (Z'). Figure reproduced from Wright (1934).

Not only did Wright hypothesize a biochemical basis for dominant phenotypes, but he appreciated the nonlinear relationship (hyperbolic response) between enzyme activity (gene dosage) and phenotype. The work of Wright illustrated that doubling the quantity of a catalyst produces less than double the amount of product following a hyperbolic correspondence (Wright, 1934). With this general gene-to-product mechanism, Wright demonstrated that heterozygote gene dosage will typically not result in intermediate phenotypes (Figure 2.2).

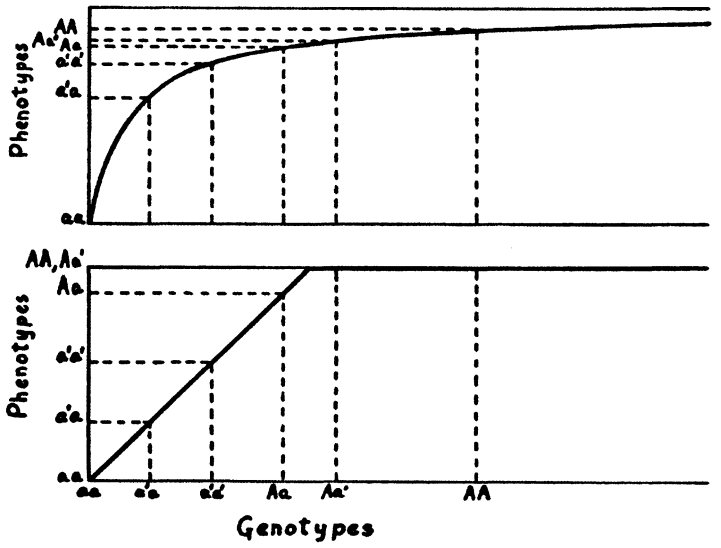


Figure 2.2
Wright represents the correlation between genotype with varying degrees of enzyme activation and corresponding phenotype predicted by his model. He represents two extreme cases of his mathematical theory: above, if phenotypic effect varies with both substrate and enzyme activity ("theory of joint control"); and substrate or enzyme is limiting (*e.g.* if catalyst is in excess, then no effect is gained through increases in catalyst efficacy or concentration) ("theory of limiting factors"). Figure reproduced from Wright (1934).

The correlations in Figure 2.2, Wright insists, are "extreme cases of a single theory" (Wright, 1934). The modeled levels of dominance were consistent with the equivalent or incomplete dominance seen in heterozygotes for various traits in organisms.

Wright's theoretical description linked, for the first time, phenotype to genotype using a biochemical and not phenomenological mechanism. Further, it provided an explanation for mutations that reversed recessive phenotypes (a phenomenon previous models could not explain) because it allowed that enzymes could be reactivated in addition to being inactivated by mutations. Although the question of dominance was initially applied to diploid systems, the answer revealed the influence of dosage of genes and hence gene products (enzymes) on the phenotype of all organisms.

Fisher, Haldane and Wright's theories, though at times antagonistic, provided a constructive paradigm shift by applying mathematical descriptions to the evolution of dominance and relating catalyzed reactions to phenotype. The integration of Darwinism and Mendelian genetics within a mathematical framework for studying the process of evolution in a quantitative manner came to be known as the modern synthesis (Huxley, 1942).

2.2 Mechanisms for dominance: Kacser & Burns and metabolic control analysis

According to Wright and Haldane, dominance is a product of enzyme function. Specifically, Wright proposed that a wild type allele confers a level of catalytic activity such that limiting the dosage of the enzyme by half has little effect on the overall rate of product formation. However, this early attempt to correlate dominance to genotype using physiological principles was restricted to properties of a single allele. In *The Molecular Basis of Dominance*, Kacser & Burns (1981) extended the early physiological mechanism put forth by Wright to explain dominance and showed "that the recessivity of mutants is an inevitable consequence of the kinetic properties of enzyme-catalyzed pathways and that no other explanation is required." The mathematical groundwork for calculating the steady-state flux through an enzyme network would be used to support this claim and would come to be recognized as seminal works in the field of Metabolic Control Analysis (Kacser & Burns, 1973; Heinrich & Rapoport, 1974a, b).

Kacser & Burns (1973) described the response (change in flux) of a metabolic pathway to modulation in effective activity of a single enzyme in terms of the enzyme's 'flux control coefficient'. Flux and the response of flux to catalytic change are both systemic properties dependent on all enzymes, which are themselves connected by shared

metabolites, e.g. the product of one is a possible substrate for another. The flux control coefficient for an enzyme can have a value between 0 (no control) and 1 (fully controlling, a metabolic 'bottleneck').[3] It is then shown that the sum of all interrelated flux control coefficients within a system is equal to unity for systems of any structural complexity (Kacser & Burns, 1973, 1979).[4] This property illustrates the limits of control attributable to any one enzyme amongst the few or many that constitute a metabolic pathway. Because control is distributed across all enzymes in a metabolic pathway only one enzyme can be said to be predominantly controlling (as the control increases for one, the control of the remaining enzymes must decrease). The trivial solution for a pathway with a single enzyme shows flux varies linearly with effective enzyme activity (Figure 2.3). The inclusion of enzymes in a linear pathway (illustrated in Equation 2-1) creates a hyperbolic response similar to that predicted by Wright (1934), and nonlinearity increases with the number of links (n) between the starting substrate (X_1) and the final product (X_n).

$$X_1 \underset{E_1}{\longleftrightarrow} S_1 \underset{E_2}{\longleftrightarrow} S_2 \ldots S_{n-1} \underset{E_n}{\longleftrightarrow} X_n \qquad 2\text{-}1$$

[3] Kacser & Burns recommend using these descriptors with caution. The flux control coefficient represents a continuum of responses, and it is imprecise to use terms like 'bottleneck' when they represent a range of control with no defined lower bounds.

[4] This includes linear or branched pathways constituted by enzymes (that undergo saturation or not) whose activities may be modulated by various interactions (activation or inhibition).

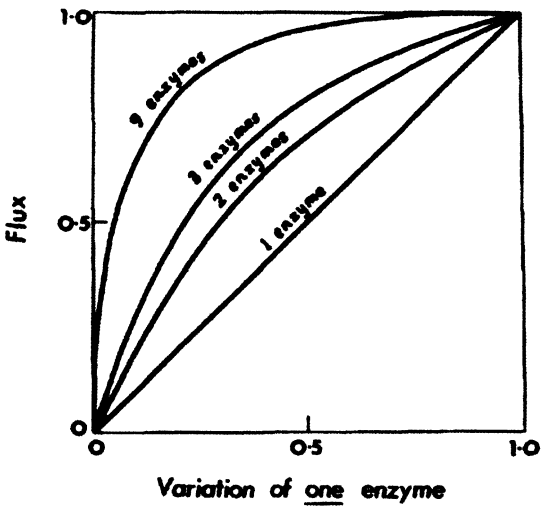


Figure 2.3
The response of flux to variation in activity of a single enzyme is dependent on the number of enzymes with which the enzyme shares a pathway and the structural organization of said pathway. Assuming the flux control coefficient is shared equally across participating enzymes, the flux control coefficient for any enzyme, which defines the response curve, scales with the number of enzymes in a series (see Equation 2-1). Figure reproduced from Kacser & Burns (1981).

The modulation of dosage of most enzymes will have little effect on flux, which is consistent with, and is the biochemical basis for, Fisher's observation that most heterozygotes present a wild type phenotype. The model also allows for intermediacy of phenotypes due to the linkage dependent nonlinearity (Figure 2.4).

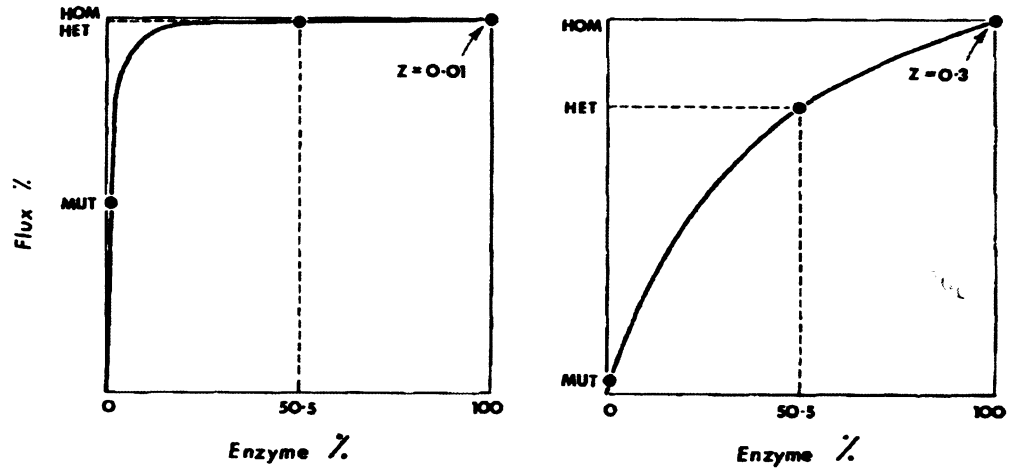


Figure 2.4
The flux control coefficient (Z) determines the response of the system to changes in effective enzyme activity.[5] Hence, the greater the flux control coefficient, the greater the change in flux relative to the change in enzyme activity. Diagrams at the left (Z=0.01) and right (Z=0.3) illustrate the change in flux for genotypes and corresponding enzyme activities relative to wild type: homozygous mutants (MUT, 1%), heterozygotes (HET, 50.5%) and homozygous wild type (HOM, 100%). Figure is reproduced from Kacser & Burns (1981).

The history of the study of dominance, from Mendel up to Kacser & Burns, relied primarily on observations of diploid organisms where changes in phenotype were attributed to gene dosage and the subsequent dosage of gene product. However, the work of Kacser & Burns definitively showed that the phenomenon of dominance arose from differences in effective enzyme activity, which by that time was known to occur by various molecular mechanisms: changes in the kinetic parameters (k_{cat} and K_M), altered expression levels, or changes in concentration of interacting metabolites that affect the reaction equilibrium at the metabolic link of interest. Thus, this mathematical 'system' centric approach is broadly applicable to the study of altered phenotypes resulting from modulation of enzyme activity due to mutation or other means in any organism.

[5] The flux control coefficient, referred to as Sensitivity and symbolized by Z in the original publications of Kacser & Burns (1973, 1981), has since been reassigned to the symbol C_i^J.

2.3 Experimental tests of relation between flux and fitness

It was predicted that growth rates are intrinsically correlated to the enzyme reaction rates, and the relation between the two follows a hyperbolic curve of diminishing returns as was seen for flux (Kacser & Beeby, 1984). Resource-limited chemostat studies confirmed that growth rates (fitness) were dependent on enzyme activities (Hartl *et al.*, 1985; Dean *et al.*, 1986). For a series of β-galactosidase mutants in isogenic *E. coli* strains grown in conditions where lactose is the limiting nutrient to growth, the fitness follows a hyperbolic curve with β-galactosidase activity (see Figure 2.5A). The large fitness cost of less active mutants would not be observed if alternative carbon sources were available. This is an example of the fitness-activity relation exhibiting selection dependence. Hartl and colleagues postulated that environmental pressures that alter the fitness-activity relation could favor increased activity that would result in negligible fitness benefit in nonselective environments for enzymes already situated on the fitness-activity plateau (Hartl *et al.*, 1985) (Figure 2.5B).

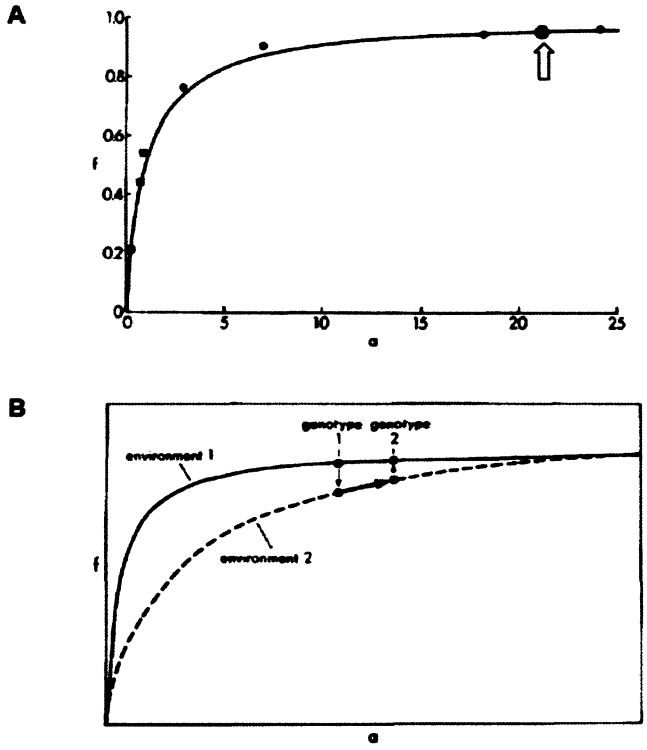


Figure 2.5
The dependence of growth rates (f) of *E. coli* on β-galactosidase activity (a) is shown to follow a hyperbolic relation, which is accentuated under culturing conditions of lactose limited media. (A) The activities of wild type β-galactosidase (circled data point) and mutants plotted against growth rates of the respective genes introduced into isogenic *E. coli* K12 followed a saturation curve. (B) The fitness-activity curves are dependent on environmental conditions where an enzyme's unique function is targeted as they were for β-galactosidase where lactose is the only carbon source. Here, genotypes 1 and 2 are hypothetical genotypes. Genotype 2 has a growth rate advantage only under selective conditions of environment 2. Both figures have been reproduced from Hartl, Dykhuizen and Dean (1985).

The study relating growth rate to the activity of a single enzymatic link, β-galactosidase, was extended to include an additional link, β-galactoside permease, forming a linear pathway of lactose import and subsequent catabolism (Dykhuizen *et al.*, 1987). In this two enzyme system, they demonstrated that the Kacser & Burns summation theory, which states that the sum of flux control coefficients is constant and equal to

unity, is true and predicts a larger control coefficient for permease. The large control coefficient would make permease the target of adaptation if the system underwent selection in a continuous culture. In general, an enzyme associated with the largest flux control coefficient can be said to the weakest link within a metabolic system. Improvements to this weak link would return the largest increase of flux and fitness.

The work of Dean, Dykhuizen and Hartl confirmed that enzyme activity was correlated to fitness with the strongest dependence occurring under targeted selection conditions. Using chemostats, they were able to differentiate small fitness differences using a strong selective pressure over many generations. In addition to this, they demonstrated the strength of using isogenic prokaryote based systems (*E. coli*) to isolate fitness effects of variation at single or multiple alleles (β-galactosidase and β-galactoside permease), a now common approach used to study adaptation in microbial populations.

2.4 Molecular evolution studied in populations of microorganisms

Microorganisms are well suited for studying evolution in the laboratory. Bacteria (Lenski & Travisano, 1994; Cooper *et al.*, 2003; Elena & Lenski, 2003; Barrick *et al.*, 2009) and viruses (Bull & Wichman, 1998; Burch & Chao, 1999; Turner *et al.*, 1999; Wichman *et al.*, 1999; Bull, 2003; Knies *et al.*, 2006) are excellent model systems for studying the interplay of environment and selection on evolutionary outcomes. The short generation time of bacteria makes it possible to observe hundreds of generations over weeks and thousands within a month. Meanwhile, near complete control over experimental growth conditions allows researchers to investigate the effect of single variables (e.g. temperature or resources) on the evolution of a population (Elena & Lenski, 2003). Snapshots of the population may be captured by intermittently sampling

from the population and preserved. These samples act as a 'fossil record' of adaptive processes for future studies. In the absence of conjugation factors, the asexual reproduction of bacteria further simplifies analysis.

Mistakes in DNA replication, recombination and repair generate mutations that precede selection. However, it is the selection that determines the fitness of the mutant. The fitness of an evolved microbial strain relative to the founding strain can be measured through head-to-head or Malthusian competitions (Elena & Lenski, 2003). Fitness, in turn, is used to define an adaptation as any heritable trait that increases the growth rate, a proxy for fitness, of an individual relative to its predecessor (see Orr (2009) for a review).

The extent to which chance, adaptation, and history are determinants for evolution was addressed succinctly with a thought experiment proposed by Stephen J. Gould. Gould suggested that by "replaying life's tape" with identical evolution experiments, the role of natural selection as the invisible hand of evolution would be observed as similar evolutionary trajectories leading to similar evolutionary outcomes (1989). Continuous evolution experiments with *Escherichia coli* populations have been propagated successfully for tens of thousands of generations with identical conditions (Bennett *et al.*, 1990; Travisano *et al.*, 1995; Cooper *et al.*, 2003). General conclusions garnered from such studies suggest that adaptation can occur over long timescales and that these evolutionary changes are in fact reproducible (Barrick *et al.*, 2009).

The full potential of adaption is not realized immediately. Continued adaptation of a population of microorganisms in a constant environment is due to the relation between fitness benefit and the length of time it takes the beneficial mutation to attain a majority

stake within the population. The smaller the difference between the fitness of two genotypes, the longer it will take for the fitter genotype to out-reproduce the less fit genotype to attain prominence. Mutations that increase fitness substantially are rare, and most mutations will be deleterious or have little or no effect (Orr, 2003). In experimental evolution studies this phenomenon has been observed as an initially rapid increase in fitness followed by increasingly smaller steps (adaptation traverses up the hyperbolic fitness-activity curve and quickly encounter diminishing returns). A change in fitness can often result from many mutations throughout the genome, some of which may have little direct influence on fitness while others may be critically important to adaptation. In a limited number of cases, microarray analysis or rational analysis of likely molecular mechanisms has been successfully used to identify sites of mutations. However, with the sharp decrease in the cost of sequencing it is no longer as challenging to find mutations as it is to link the mutations to their respective phenotypic effects.

Focusing the selective pressure to a specific enzyme function, similar to the targeted selection used by Dean *et al.* (1986), simplifies the search for mutations by favoring adaption of the gene responsible for this unique function and perhaps genes responsible for its regulation. The targeted selection reduces the complexity of the fitness landscape by limiting the size of the underlying sequence space (a gene versus the entire genome). As mutations arise, evolving genes traverse such fitness landscapes and are continuously under selection to be adept at a specific biochemical function while minimizing metabolic costs to the cell.

Many mutational pathways exist between the founding and final sequence in a population. While the founding strain is predetermined and the final fixed strain or strains

can be easily determined, it is the intermediates that are critical to understanding why the selection favored a particular pathway. The evolutionary pathways or trajectories across this complex fitness landscape are much more difficult to identify, and yet are essential to prediction (Hall, 2002).

2.5 Chapter summary

The biochemical basis for fitness was conceived by Wright (1934) and formalized by Kacser & Burns (1981). The theory of metabolic control predicts that the structure of metabolic pathways produces an inherent robustness to mutations that alter enzyme activity. However, the use of targeted selection conditions allowed Dean, Dykhuïzen and Hartl (1985) to correlate fitness to enzyme activity. Evolutionary biologist, meanwhile, were studying evolutionary phenomena such as dynamics and genetic bases of adaptation in microbial populations.

Dean and Thornton noted the convergence of these approaches and coined the term 'functional synthesis' to capture the synergy between evolutionary and molecular biology to address important questions such as the evolution of complexity and antibiotic resistance (Dean & Thornton, 2007). The functional synthesis, in its most fully realized form, is an integrated systems biology approach to evolutionary dynamics that links physicochemical properties of molecules to evolutionary fates in a quantitative and predictive manner.

Functional synthesis flourishes in an experimental framework that allows investigators to directly link population dynamics (fitness) to changes in molecular function that result from alterations at the nucleotide level. The 'weak link' approach,

which will be discussed in the following chapter, was developed to tightly couple adaptive changes within the genome to changes in fitness and provide a population based approach that can be used to examine alterations in function and fitness at the level of atomic structure and function.

Chapter 3 – The weak link

Disruption of the interrelated chemical activities of an organism is death; by the same token, these chemical activities, collectively, are life.
Daniel E. Atkinson (1977)

Metabolism is a complex network of reactions responsible for utilizing chemical energy to synthesize components required for growth and proliferation. The enzymatic links within the metabolic network that accelerate these reactions are all, at some level, regulated by their respective substrates and products or by products of reactions in separate pathways. Given that the reactions are all contained within the intracellular milieu, it can be said that no enzyme functions in isolation. It is due to this interconnectedness of metabolism that is has been difficult to attribute the genetic variation in genes, let alone kinetic properties of a single enzyme, to the phenotype of the entire organism. This has come to be known as the genotype-to-phenotype problem. Historically, additional complications arise from a dearth of the genetic data needed to compare the genomes of phenotypic variants and phenotypes that are hard to describe quantitatively. The weak link approach reduces these limitations by favoring adaptation to a single locus through natural selection with the dual benefit of constraining the region of influential variation to a smaller genetic landscape and emphasizing quantifiable differences in fitness.

The weak link strategy requires an essential gene that is maladapted in relation to the host system and a selective pressure that will favor improvements to this weak link. Couñago *et al.* created a weak link using a homologous recombination strategy to replace the chromosomal copy of the essential adenylate kinase gene (*adk*) of the thermophilic

bacterium *Geobacillus stearothermophilus* with an ortholog from the mesophile *Bacillus subtilis* (2005). Recombinant *G. stearothermophilus* cells that expressed only *B. subtilis* adenylate kinase (AK_{BSUB}) were unable to grow at temperatures higher than 55 °C due to heat inactivation of the mesophilic enzyme and consequent disruption of adenylate homeostasis (Couñago & Shamoo, 2005). Continuously growing populations of the recombinant bacteria were then subjected to selection at increasing temperatures (from 55 to 70 °C) that favored changes in the one gene not adapted for thermostability, *adk*. During the course of selection, the population was sampled and intermediates of adaptation were observed as mutations to *adk* and the corresponding proteins were examined to determine how the mutant populations traversed the adaptive landscape to increased fitness (Couñago *et al.*, 2006).

The weak link method is useful because it bypasses many of the problems in tying genotype to phenotype that have thwarted previous experiments, specifically by subjecting a single gene to an environmental pressure. By choosing adenylate kinase, an essential enzyme central to energy metabolism, enzyme performance would be closely tied to cellular viability. Evolution at the locus would arise within the population due to differences in survivability or growth rate (fitness) between the starting strain and mutants, properties that are easily quantified. To understand why variation at the level of the enzyme, adenylate kinase, will be reflected in organismal viability, one must begin with the central role of ATP in energy metabolism and follow with the unique role of AK in maintaining ATP reconversion.

3.1 ATP, the basis of the cellular energy economy

There are three fundamental requirements for life as perpetuated by metabolic systems within cells: energy in the form of ATP, a reducing potential provided by NADPH or NADH, and raw materials used in biosynthesis. This is not to say that ATP and nicotinamide nucleotides are the only essential coupling agents required for life, but they are unique in that their roles are coupled across the metabolic pathways associated with catabolism and anabolism. Catabolic pathways such as glycolysis and the citric acid cycle increase ATP and NADH concentrations far above their equilibrium states. The chemical energy conserved from degradative processes in the form of the non-equilibrium states is used by biosynthetic pathways to form complex macromolecules on which the cell depends (anabolism).

The intracellular milieu abounds with thousands of enzymes and chemical compounds interacting, reacting and regulating to form a chorus of reactions to produce what we call metabolism. Activating carriers such as ATP increase the overall equilibrium constant of many such reactions and series of reactions by increasing the equilibrium ratios for each metabolic step without needing high concentrations of metabolic intermediates to drive the reaction. I will illustrate the power of activated phosphate carriers with a simple reaction where a phosphate from ATP is transferred to glucose to form glucose 6-phosphate (glucose 6-P). Reaction 3-1 is endergonic, meaning the Gibbs free energy ($\Delta G'°$) is positive and by definition the reaction is not thermodynamically spontaneous under standard conditions.

$$\text{glucose} + P_i \rightleftharpoons \text{glucose 6-P} + H_2O \qquad \text{3-1}$$
$$\Delta G'° = 13.8 \text{ kJ/mol}; \; K'_{eq} = 3.9 \times 10^{-3} \text{ M}^{-1}$$

At equilibrium, the concentration of inorganic phosphate (P_i) required to obtain a tenfold enrichment in the concentration of glucose 6-P relative to glucose is approximately 2500 M, which is a physically absurd value. However, the phosphorylation of glucose is the first, preparatory step of glycolysis and must be achievable in cellular conditions. Enzymes (e.g. hexokinase) cannot shift the equilibrium determined by the free-energy change of the reaction, and only accelerate the process by which the equilibrium is reached. Activated carriers are capable of changing the equilibrium and driving unfavorable reactions forward by coupling an endergonic reaction such as the phosphorylation of glucose to the exergonic hydrolysis of ATP (reactions 3-2 and 3-3).

$$\text{ATP} + H_2O \leftrightarrows \text{ADP} + P_i \qquad 3\text{-}2$$
$$\Delta G'^{\circ} = -30.5 \text{ kJ/mol}; K'_{eq} = 2.0 \times 10^5 \text{ M}$$

$$\text{ATP} + \text{glucose} \leftrightarrows \text{ADP} + \text{glucose 6-P} \qquad 3\text{-}3$$
$$\Delta G'^{\circ} = -16.7 \text{ kJ/mol}; K'_{eq} = 7.8 \times 10^2$$

Coupling the large, negative, standard free energy of hydrolysis of ATP to the endergonic phosphorylation of glucose can increase the enrichment of glucose 6-phosphate by 200,000 fold under standard equilibrium conditions without needing greater concentrations of phosphate.[6] Because metabolic reactions can be made more favorable at low concentrations of intermediates through activating agents such as ATP, cells have a greater capacity for pathways in metabolism.

[6] The example provided is accurate under standard conditions (25 °C and 1 M concentrations) that cannot be obtained in living cells due to differences in temperature, concentrations of adenylate moieties and phosphate, and the binding of magnesium, which all affect measured free energies.

The Gibbs free energy is an extensive property, meaning its value varies with the amount within a thermodynamic system.[7] The effectiveness of ATP as an activated carrier, also known as its phosphorylation potential (ΔG_p), is dependent on the concentration within the system. As the concentration of ATP increases relative to ADP, the free energy of hydrolysis becomes a larger, negative value (ln(x) as x decreases).

$$\Delta G_p = \Delta G'^{o} + RT \ln \frac{[\mathrm{ADP}][\mathrm{P_i}]}{[\mathrm{ATP}]} \qquad 3\text{-}4$$

Given Equation 3-4 which relates the actual free energy of hydrolysis to temperature, reactant and product concentrations, a decrease in the ratio of ADP/ATP to 0.1 (less ADP) or increase to 10 (more ADP) corresponds to a decrease in free energy (ΔG_p = 30.5 kJ/mol − 2.3 kJ/mol) or an increase in actual free energy (ΔG_p = 30.5 kJ/mol + 2.3 kJ/mol), respectively. Thus, the spontaneity of reactions coupled to the hydrolysis of ATP is dependent on its relative concentration. As the ATP concentration increases, there are more potential phosphate donors and the free energy for the transfer increases which makes these biosynthetic reactions more thermodynamically favorable. It is no wonder that natural selection has conserved metabolic pathways responsible for the maintenance of intracellular concentrations of ATP far above hydrolysis equilibrium.

ATP is thermodynamically unstable due to charge repulsion and resonance stabilization of the leaving phosphate group, hence the large, negative free energy of

[7] The standard free energy ($\Delta G'^{\circ}$) is a constant, whereas the actual free energy (ΔG), which determines the spontaneity of the reaction, is a function of both reactant and product concentrations as well as temperature by the following formula: $\Delta G = \Delta G'^{o} + RT \ln \frac{[C]^c[D]^d}{[A]^a[B]^b}$ where R is the gas constant (8.314 kJ/mol), T is the absolute temperature, and *a*, *b*, *c* and *d* represent the stoichiometry of reactants *A* and *B* and products *C* and *D*.

hydrolysis. However, ATP is kinetically stable meaning that, despite an equilibrium that lies towards a hydrolyzed state, the rate at which equilibrium is reached in solution is very slow when uncatalyzed. Likewise, ATP is slow to react with other chemicals at the expense of the large activation energy ($\Delta G^{\ddagger}$ = 200-400 kJ/mol) required to break phosphoanhydride bonds (Lehninger *et al.*, 2008). Cells utilize the thermodynamically unstable / kinetically stable duality to direct chemical reactions using the catalytic properties of enzymes.

ATP not only drives biosynthetic reactions, but it and its fellow adenylate moieties are the biosynthetic components for various molecules and polymers. ADP is a precursor for the deoxy form that is incorporated into DNA, and ATP is one of four ribonucleoside 5'-triphosphates used in the construction of RNA. In protein synthesis, ATP is required for tRNA aminoacylation and contributes a nitrogen and carbon from its purine ring in the synthesis of the amino acid histidine. A summary of the interconversions of adenylate moieties through formation and utilization are represented in Figure 3.1.

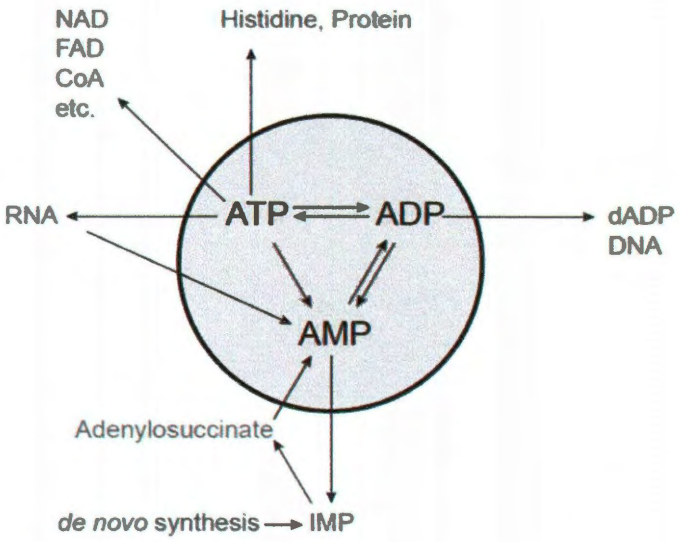


Figure 3.1
Arrows contained within the circle represent the interconversion between adenylate species attributed to ATP-utilizing or ATP-regenerating pathways and the adenylate kinase homeostasis reaction. These reactions do not alter the total concentration of adenine nucleotides. Arrows that pass in and out of the circle identify common utilizations of adenine nucleotides as precursors of macromolecules and cofactors which is balanced against the synthesis of AMP and recovery from unstable RNA. This figure is adapted from a review by Chapman and Atkinson (1977).

3.1 Adenylate homeostasis

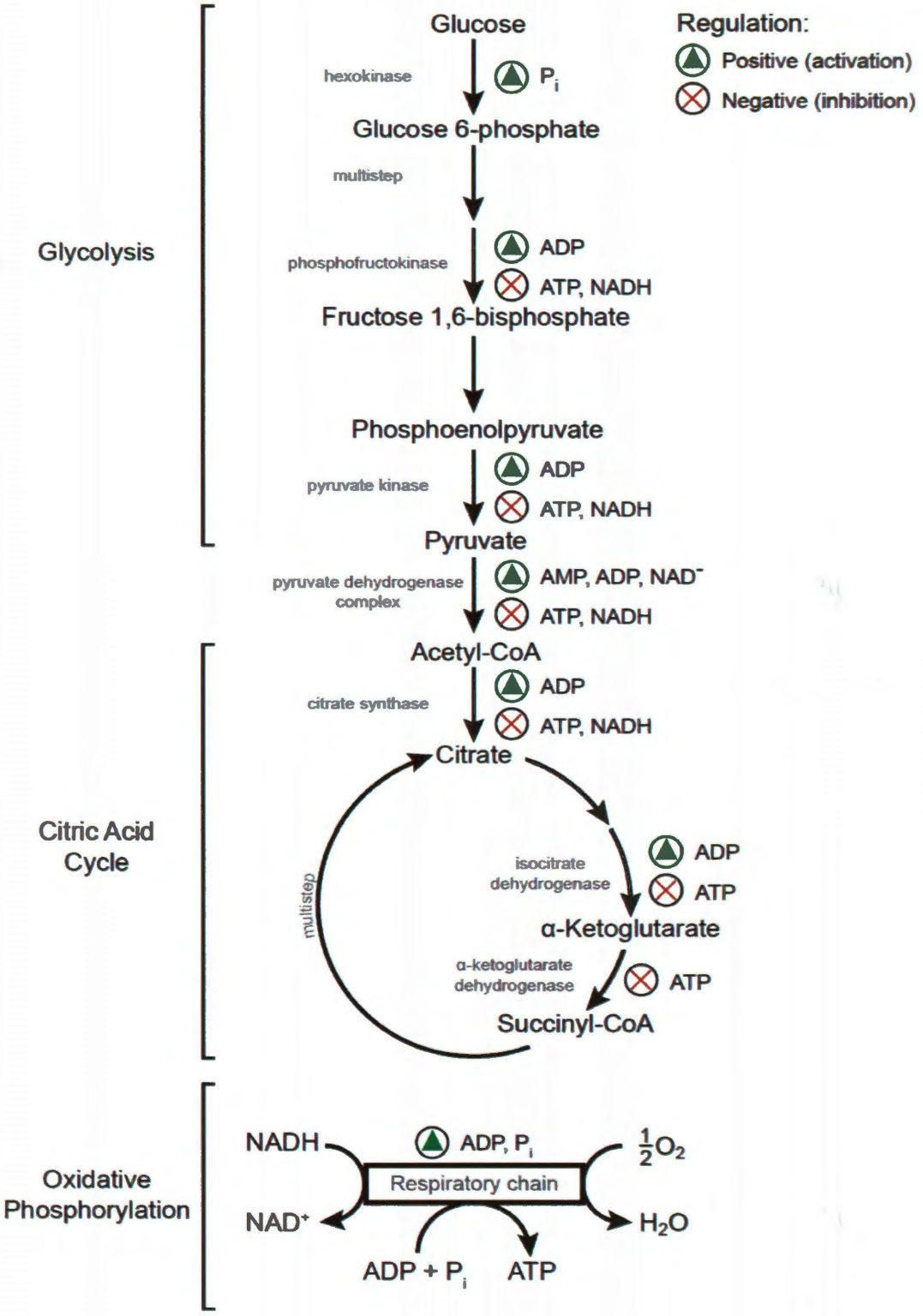
The free energy of ATP hydrolysis that drives endergonic reactions is higher when the [ATP]/[ADP] ratio is high. The energy available to the cell for a given [ATP] is lower when the ratio falls and greater when it rises. For this reason it is essential for cells to maintain a high concentration of ATP, and evolution has produced mechanisms to maintain the phosphorylation potential under various conditions.

Cells capable of aerobic respiration regenerate ATP by way of glycolysis, citric acid cycle and oxidative phosphorylation. The phosphorylation of glucose, described earlier, is the first step of glycolysis, which ultimately yields two ATP molecules for each

glucose molecule degraded into pyruvate. The citric acid cycle and oxidative phosphorylation further extract energy from the remaining pyruvate for aerobic organisms. Most ATP is generated by oxidative phosphorylation, specifically by ATP synthase which is powered by a proton gradient. The proton gradient is generated by the flow of electrons in the respiratory chain via NADH and $FADH_2$ produced by the citric acid cycle and the oxidation of fatty acids. The proton motive potential is the transformation of free energy captured by electron transport in respiring cells. ATP synthase couples the translocation of protons across a gradient to the endergonic phosphorylation of ADP to regenerate ATP, the chemiosmotic paradigm first postulated by Peter Mitchell (1961).

ATP-producing pathways are all regulated at the level of enzymatic reactions by either adenylate moieties positively (AMP, ADP or P_i) or negatively (ATP or NADH) or by product feedback to match throughput with the needs of the cell. Hans A. Krebs was the first to suggest that the predominance of glycolysis over gluconeogenesis was regulated by AMP concentration (1964). Many such regulatory interactions have been identified and are summarized extensively by Daniel E. Atkinson and are detailed in Figure 3.2 (Atkinson, 1966, 1969; Atkinson, 1977).

Figure 3.2
Control points along ATP-regenerating pathways are sensitive to AMP, ADP (adenine nucleotide species not utilized for chemical work) and ATP (end product). Regulation of these pathways increases the recovery of ATP from ADP when the cell experiences too little ATP, which is determined by a low relative ratio of ATP/AMP, ATP/ADP or NAD^+/NADH dependent on the enzyme. Inhibition (red X) or stimulation (green triangle) can adjust flux through the pathways to ensure energy needs are met. Figure adapted from Lehninger (2008). Additional metabolites that play a regulatory role have been ignored to focus on ATP, NADH and their moieties.


Glycolysis
Glucose
hexokinase
P_i
Glucose 6-phosphate
multistep
phosphofructokinase
ADP
ATP, NADH
Fructose 1,6-bisphosphate
Phosphoenolpyruvate
pyruvate kinase
ADP
ATP, NADH
Pyruvate
pyruvate dehydrogenase complex
AMP, ADP, NAD⁻
ATP, NADH
Acetyl-CoA
Citric Acid Cycle
citrate synthase
ADP
ATP, NADH
Citrate
isocitrate dehydrogenase
ADP
ATP
α-Ketoglutarate
α-ketoglutarate dehydrogenase
ATP
Succinyl-CoA
multistep
Regulation:
Positive (activation)
Negative (inhibition)
Oxidative Phosphorylation
NADH
ADP, P_i
$\frac{1}{2}O_2$
Respiratory chain
NAD^+
H_2O
ADP + P_i
ATP

The "adenylate control hypothesis" postulates that the series of enzymes that make up the ATP-regenerating sequences respond, not to the concentration of a single metabolite, but to cell's momentary concentrations or ratios of adenylate moieties (Atkinson, 1966). To account for the heterogeneity of responses of various enzymes to concentrations of individual adenylates or their ratios (ATP/ADP or ATP/AMP), a single parameter was defined to be applied to the study of metabolic enzymes that fall under adenylate regulatory control: energy charge (Equation 3-5) (Atkinson & Walton, 1967).

$$\text{Energy charge} = \frac{[\text{ATP}] + 1/2[\text{ADP}]}{[\text{AMP}] + [\text{ADP}] + [\text{ATP}]} \qquad 3\text{-}5$$

In essence, energy charge is an expression of the molar ratio of ATP and half the molar ratio of ADP (the adenylates with phosphoanhydride bonds). Energy charge as a fundamental metabolic control parameter can vary between values of 0 and 1 for a discharged state (only AMP) and fully charged state (only ATP), respectively.[8] The parameter provides a general description for all adenylate control of ATP-regenerating and utilizing pathways to maintain an adenylate balance according to the needs of the cell. A range for energy charge values in metabolizing cells is reported to be 0.87 to 0.94 (Atkinson, 1977).[9] It was later shown *in vivo* that a decrease in energy charge for *E. coli* during starvation conditions corresponded to a decrease in viability of the cells (Chapman *et al.*, 1971). This work suggested that if energy charge could no longer be maintained by metabolic regulation, the low ATP concentration would result in a "lethal disintegration of the cellular economy" (Chapman *et al.*, 1971).

---

[8] Atkinson admits "modulation of enzymic behavior cannot, of course, depend directly upon an abstract concept like energy charge. The enzyme must respond to the concentration of one of the nucleotides or to some function of the concentrations of two or all three" (1968).

[9] Chapman and coworkers compiled values for energy charge across various eukaryotic and prokaryotic organisms and tissues at different metabolic states (1971).

Energy charge is a more robust parameter to compare intracellular energy balance than ATP concentration because it represents a ratio of concentrations and is not susceptible to limitations that produce discrepancies in the latter (Chapman & Atkinson, 1977). An accurate value of total ATP within a living cell is hard to determine because (1) current methods for quantifying total ATP destroy the *in vivo* state by disrupting the cell membrane, (2) the value is highly sensitive to methodological variation, and (3) a fraction of all ATP is sequestered by association with other macromolecules. No consensus has been reached on the either the value of total ATP or the means of measuring it. A value of ~3 mM has been reported (Bochner & Ames, 1982) and echoed within the literature (Buckstein *et al.*, 2008). More recently, luciferase based assays have been used to determine available ATP *in vivo* (Schneider & Gourse, 2004) while a fluorescent sensor has revealed ratios of ATP to ADP with temporal and spatial resolution *in vivo* (Berg *et al.*, 2009). The value of "total ATP" may however have little relevance to metabolic regulation or "growth rate-dependent control" of rRNA transcription as described by Schneider and Gourse (2004). In light of the adenylate control hypothesis, the high energy charge equilibrium is maintained by co-regulating ATP-utilizing and regenerating reactions which are regulated in a competing manner by different adenine nucleotides. Thus, the competition between adenine nucleotides supports the need of a parameter that takes into account ratios (energy charge) over the absolute concentration of a single agent.

3.2 Adenylate kinase, an essential link in energy metabolism

In addition to the complex regulatory role of adenylate moieties, adenylate kinase provides an essential role in the interconversion of the three adenine nucleotides. The

ATP-regeneration pathways such as oxidative phosphorylation require a pool of ADP with which to recharge to ATP. Adenylate kinase (EC number 2.7.4.3) is an ATP:AMP phosphotransferase that carries out a reversible reaction (Equation 3-6) essential for maintaining a pool of ADP for ATP-regenerating pathways (at high energy charge the AK reaction favors ADP production).

$$ATP + AMP \leftrightarrows 2ADP \qquad 3\text{-}6$$

At first this reaction may seem counterproductive given the system's propensity for requiring and generating ATP, but the phosphoryl transfer effectively pulls AMP, which does not contribute to energy charge, up into a state where it can be recharged by the ATP-regenerating pathways. Thus, the essential role of AK in adenylate homeostasis is to rescue AMP generated *ab initio* or metabolically at the cost of an ATP to provide a substrate for the cell's ATP-regenerating machinery.[10]

Heat-sensitive *E. coli* strains (CR341T28 and CV15) identified through independent studies that contain thermosensitive adenylate kinase demonstrated a positive correlation of AK activity to viability and energy charge (Cousin & Belaich, 1966; Glaser *et al.*, 1975; Glembotski *et al.*, 1981). Studies showed that even a transient excursion to non-permissible temperatures resulted in sudden decrease in adenylate energy charge (0.9 to 0.2) that was concomitant with cell death (Glembotski *et al.*, 1981). Inactivation of AK at nonpermissive temperatures resulted in a joint decrease in ATP and

[10] Athel Cornish-Bowden has written of a metabolic role for AK postulating that the interconversion has a greater impact on AMP concentration given intracellular conditions and thermodynamics of the reaction (1979). Despite small changes in energy charge, fractional AMP concentration can vary much more greatly, and the corresponding changes can have dramatic effects on enzymes regulated by AMP.

increase in AMP concentrations which in turn was correlated to a decline in protein, DNA, RNA and phospholipid biosynthesis and cell growth (Glaser *et al.*, 1973; Glaser *et al.*, 1975). The thermolability of AK in both strains would later be attributed to the substitution of a serine for proline-87 (Esmon *et al.*, 1980; Gilles *et al.*, 1986; Haase *et al.*, 1989). These studies were the first to bridge AK activity and its unique physiological role in the control of energy charge and the ability of the cell to synthesize components critical to viability and growth.

3.3 Summary of work and findings by Couñago *et al.*

Couñago and Shamoo developed a system utilizing the moderate thermophile *G. stearothermophilus* for investigating the evolution of a single gene to increased thermostability (2005; 2006). The chromosomal copy of adenylate kinase in *G. stearothermophilus* was replaced with an ortholog from the mesophile *Bacillus subtilis* (76% identity) by homologous recombination. The recombinant strain, *G. stearothermophilus* NUB3621-R:ThEv, has a temperature-sensitive phenotype due to the disruption of adenylate homeostasis at temperatures above 55 °C (Couñago & Shamoo, 2005).

Adenylate Kinase (AK) from *B. subtilis* serves well as the weak link in the evolution of thermostability as it is strongly maladapted relative to the remaining genes in the chromosome of *G. stearothermophilus*. AK is essential for maintaining adenylate homeostasis and, as with previously studied thermolabile AKs, loss of function at nonpermissive temperatures results in a lethal phenotype (Glembotski *et al.*, 1981; Couñago & Shamoo, 2005). The lethal phenotype is indicative of AK's essential role in cellular adenylate homeostasis (AMP phosphorylation to ADP), and that there are no

other enzyme or pathway complements to this function natively in *G. stearothermophilus* NUB3621-R.

The energy charge values for strains NUB3621-R and NUB3621-R:ThEv at 50 °C were very similar (0.22 ± 0.003 and 0.23 ± 0.031, respectively).[11] After growing both strains at 65 °C for 30 min, there was a marked increase in energy charge for NUB3621-R (0.45 ± 0.015) and no determinable value for the recombinant due to high adenylate concentrations in the media. Viability and energy charge of NUB3621-R:ThEv observed for cells growing at 50 °C could not be recovered once cells were returned to the permissive temperature after 20 minutes at 65 °C. This demonstrated that the metabolic state, once disrupted, cannot be restored. This conclusion is supported by Glembotski and colleagues with their work on a temperature-sensitive adenylate kinase containing strain of *E. coli* that could not be recovered after a 1.5 hour stay at nonpermissive temperatures (1981).

The previous work created a weak link in the form of a maladapted adenylate kinase and observed a clear temperature-sensitive phenotype for the strain. With this system it was possible to favor adaptations to adenylate kinase and then tie variation in the gene back to improvements in fitness at higher temperatures. To study adaptation of this weak link, Couñago *et al.* (2006) initiated a culture of NUB3621-R:ThEv that was maintained in a state of exponential growth using a turbidostat (a fermentor capable of maintaining a constant cell density). A population size of ~5 x 10^{10} cells was cultivated

[11] These values are very low relative to the range typical for metabolizing cells of eukaryotes and prokaryotes of 0.8-0.9 (Chapman and Atkinson, 1977). It is suggested by Couñago and Shamoo that this low value is due to a stressed metabolic state of cells grown at 50 °C, which is near the lower limit of their growth range (47 °C). However, the EC of NUB3621-R at a more optimal temperature (65 °C) remains low relative to the consensus range.

with an initial incubation period of 48 hr at 55 °C and then was increased daily by 0.5 °C, from the permissive 55 °C into a nonpermissive regime and up to 70 °C. Spontaneous mutations that increase fitness (increased thermostability while conserving activity) were selected for and became enriched within the population. Samples withdrawn from the turbidostat at various temperatures were plated and *adk* sequenced to quantify the frequency of mutants in the population (Figures 3.3 and 3.4). Early screening of the population at higher temperatures than those from which it was sampled can identify adaptations before they reach a majority status within the population.

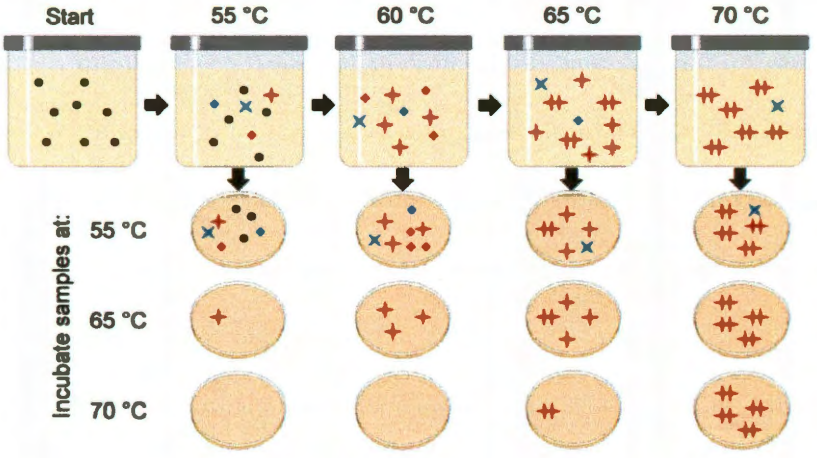


Figure 3.3
Schematic of the selection and screening for the molecular evolution project. The fermentor is inoculated with a homogenous population of the ancestral strain (NUB3621-R:ThEv, ●). Spontaneous mutations generate new strains (◆ and ✕), some of which alter *adk* and increase the fitness relative to the ancestral strain (◆ and ✚). It is important to note that this is a gross simplification, and that any of the symbolic mutations represent many possible mutations. Double mutants can occur through secondary substitutions to an existing mutant (✚✚) or in the very rare event of simultaneous mutations. The schematic represents the changing populations of strains, and the process by which the mutant lineages can be isolated prior to taking a foothold within the population.

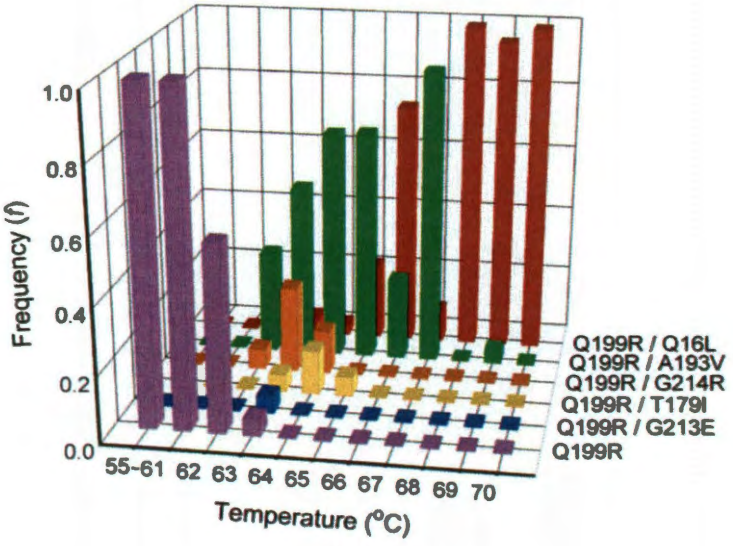


Figure 3.4
Frequency of adenylate kinase mutations fluctuate in a temperature dependent manner. The *adk* gene was sequenced in 242 isolates taken from samples pulled at one degree intervals across the range of temperatures. Substitution Q199R was fixed at 55 °C and served as a progenitor for the mutations that followed. The decrease in frequency of Q199R corresponded with the appearance of mutants with an additional mutation within the *adk* gene. The double mutant genotypes will decline in frequency with increasing temperature with the exception of Q199R/Q16L.

The strengths of this experimental evolution system are the breadth of the sequence space explored through spontaneous mutations due in part to the large continuous population size and the ability to retrace the mutational events that constitute the adaptive walk. In the original study, six mutants were isolated from the evolving population: a single mutant and five double mutants (Figure 3.5 has the mutations mapped onto the protein structure). A mutation was found in *adk* for all 242 isolates. Of the six adaptive mutations to *B. subtilis adk*, five are transitions, meaning the nucleotide changes but the purine or pyrimidine nature is conserved, and one is a transversion, where a purine changes to pyrimidine or vice versa. The CAG→CTG transversion is

responsible for glutamine-16 to leucine substitution (Q16L). G214R can be created by either transition (GGA→AGA) or transversion (GGA→CGA). Despite this redundancy, the transversion was never observed in the ten sequences of from the G214R containing isolates. These two observations provide circumstantial evidence supporting transition bias as a force shaping evolutionary dynamics in the weak link system. Due to the strong selective pressure, no "neutral" mutations were observed that altered the encoding nucleotide sequence but not the translated primary sequence.

Figure 3.5
Structure of AK_{BSUB} Q199R (purple ribbons for the mainchain) complexed to the transition state inhibitor Ap_5A (sticks) (PDB ID: 2eu8). The positions of the Cα atoms of the five adaptive mutations that conferred additional fitness at higher temperatures are highlighted as: AK_{BSUB} Q199R/Q16L, red (PDB ID: 2osb); AK_{BSUB} Q199R/T179I (PDB ID: 2oo7), yellow; AK_{BSUB} Q199R/A193V, green (PDB ID: 2ori); AK_{BSUB} Q199R/G213E, blue (PDB ID: 2qaj); and AK_{BSUB} Q199R/G214R, orange (PDB ID: 2p3s).

Each mutant's rise and fall within the population was identified through isolation and subsequent DNA sequencing (Figure 3.3 and Figure 3.4). The five recovered double mutants represent a small fraction of the 1289 possible missense mutations to the single mutant progenitor. The local sequence space was well surveyed. With ~5 x 10^{10} cells in the turbidostat at any given time over approximately 1500 generations and assuming an estimated mutation frequency of ~5 x 10^{-10} (Drake *et al.*, 1998), there were approximately a million point mutations to the 651 base pair *adk* during the course of the experiment. When the weak link experiment was repeated, the same three predominant mutant strains (containing Q199R/Q16L, Q199R/T179I, and Q199R/A193V) were isolated suggesting that the evolutionary outcome is reproducible under conditions of strong selection and a large population (Couñago *et al.*, 2006).

The phenotypic influence of each mutation was characterized by recreating experimental intermediates through introduction of the mutant *adk* back into the ancestral organism by homologous gene replacement. Recombination of the mutant genes into the unadapted NUB3621-R:ThEv conferred similar growth rates to those observed for the adapted strains (Couñago *et al.*, 2006). Mutations outside of *adk* could contribute to the small growth rate advantage held by the strains isolated from the turbidostat over the unadapted strains containing the respective *adk* point mutations. However, the temperature resistance conferred by mutations to *adk* alone suggests that mutations outside of *adk* have little influence on the selective advantage of these strains.

The changes in the organismal fitness are correlated to the corresponding changes in stability and activity maxima (Figure 3.6). Stability was determined by thermal denaturation monitored by circular dichroism (CD) at 220 nm. The melting point for wild

type was 54.9 °C, and the single and double mutations were shown to confer increases of 1.7 to 13.8 °C. The amount the melting point increased loosely corresponds with the temperature at which the mutant rose to prominence within the population. Maximal activity, measured by ADP production, also shifted towards higher temperatures.

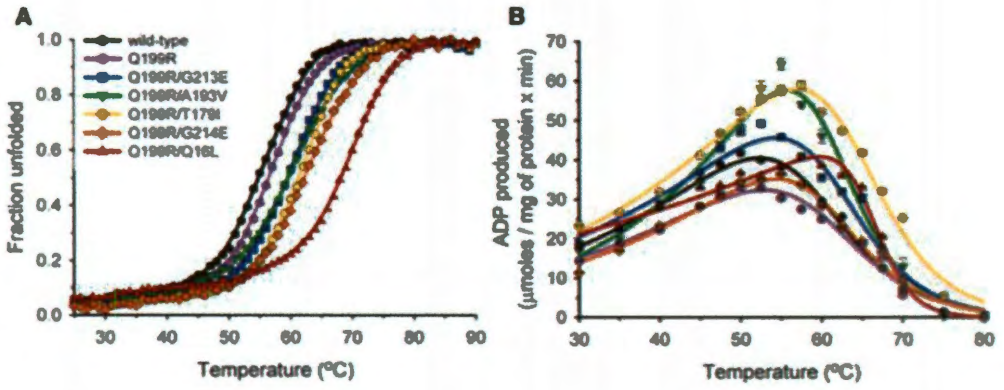


Figure 3.6
Success of mutants during experimental evolution roughly correlates to physiochemical characteristics of the isolated mutants such as stability and activity. (A) Circular dichroism is used to observe the characteristic two-state unfolding of AK as a function of temperature to quantify the stability. The data confirm expectations that the mutants would be more stable that wild type AK and that AK_{BSUB}Q199R/Q16L would be the most stable of the genotypes identified. (B) Enzymatic activity assay quantifies the production of ADP at a given substrate concentration and at various temperatures. There are noticeable changes in the activity profiles of the double mutants relative to AK_{BSUB}, either increased maximal activity (AK_{BSUB}Q199R/T179I and AK_{BSUB}Q199R/A193V) or shifts towards higher temperatures (AK_{BSUB}Q199R/Q16L). Figure is taken from (Couñago *et al.*, 2006).[12]

The mutant Q199R differs from the ancestral gene by a single mutation, and yet the mutant was broadly more successful than wild type. A modest increase in *in vitro* stability (3 °C) relative to wild type translated to an increase in the upper limit of the growth range by approximately 9 °C in the turbidostat (Couñago *et al.*, 2008). The weak

[12] The values and rank of melting temperatures (T_m) and activity profiles could not be reproduced when performed independently at a later time by CD (not shown) or by DSC (Figure 6.3A).

link study tightly coupled organismal survival to variation within a single gene, and Q199R provided for a unique opportunity to couple an adaptive advantage to structural and kinetic influences of a single mutation. Crystallographic study of the glutamine-199 to arginine mutation identified a novel ionic-interaction that may serve to stabilize the polypeptide loop that forms part of the ATP-binding site and is directly connected to the C-terminal helix (Couñago *et al.*, 2006). Site-directed mutagenesis and solvent screening studies supported the hypothesis that the formation of an ionic interaction observed in the crystal structure facilitated by the mutation stabilized the enzyme (Couñago *et al.*, 2008). Decreased unfolding rates for AK_{BSUB}Q199R relative to wild type determined by time-resolved folding experiments were also consistent with the general stabilization observed by other methods (Couñago *et al.*, 2008).

Although structure and stability are critical to enzyme function, it is the enzyme's activity that is the ultimate criterion required for maintaining adenylate homeostasis in the case of adenylate kinase, its metabolically relevant role. The rate of ADP production was determined at various temperatures (27 – 60 °C) and substrate concentrations for AK_{BSUB} and AK_{BSUB}Q199R to compare their temperature dependence of the steady-state kinetic parameters: K_M and k_{cat}. Enzymatic activity and catalytic efficiency (k_{cat}/K_M) of AK_{Q199R} was shown to be lower than wild type at temperatures below those under which the mutant was selected for (≤45 °C). However, at temperatures at and above those of the selection (≥55 °C) the catalytic activity of the mutant was increased relative to wild type with the efficiency becoming more similar. The influence of the Q199R mutation could be described as creating a stability-activity trade-off whereby rigidification of the protein through a new, stabilizing ionic-interaction inevitably limited the mutant enzyme's

catalytic proficiency at lower temperatures but promoted greater stability and thus activity at higher temperatures (Couñago et al., 2008).

3.4 Chapter Summary

The weak link method was the first successful demonstration that adaptation could be favored within a single gene using gene replacement and a targeting pressure (in this case temperature) (Couñago *et al.*, 2006). What's more, a second run of the experiment produced three of the same mutations, showing evolution as a process is not inherently random. The success of the method can be attributed to thoughtful choice of a weak link, in this case adenylate kinase, whose function is essential and thus correlated to metabolic stability. Because the viability of mutants was so tightly coupled to variation at a single locus, biochemical studies could be performed on the gene products to discern the physicochemical properties most closely correlated with a strains evolutionary success.

Initial biochemical studies of the mutants, enzyme assays to monitor activity as a function of temperature and circular dichroism to determine shifts in stability, provided a cursory description of the properties that were correlated to a mutant's success. However, a truly quantitative model relating enzyme activity and stability to fitness remained elusive. It was through much work that a molecular hypothesis for the adaptive effect of Q199R was formed, and it would require a similar focus on the kinetic properties to unravel how variation at the level of adenylate kinase gave rise to the observed evolutionary dynamics.

Chapter 4 – Enzyme activity and stability

Catalysis of chemical reactions by enzymes is essential to life because it vastly increases rates between specific interacting chemical constituents, a requisite for metabolism. As mentioned in the Chapter 2, a physiological basis for dominance was postulated in light of advancement in our understanding of enzyme catalysis. Work by Kacser & Burns, amongst others, demonstrated that enzyme activity is closely tied to fitness if the enzyme's unique function is essential under growth conditions.

In this chapter I will provide a cursory review of enzyme kinetics and stability as they pertain to my study of adenylate kinase variants and to enzymes in general.[13] As described in Chapter 3, the study of adaptation under an increasing thermal pressure introduces protein unfolding as a factor that inhibits enzyme function. The stabilization conferred by the observed mutations to adenylate kinase (e.g. Q199R/Q16L) supports the hypothesis that function can be conserved by increasing the fraction folded at higher temperatures (Couñago *et al.*, 2006), but previous studies do not speak to the altered kinetic properties (k_{cat} and K_M) of double mutants at these elevated temperatures. The work in this thesis involves the deconvolution of the temperature dependent enzyme catalysis and protein unfolding to determine which characteristics best correlate to fitness. Although unfolding is negligible in most enzymatic studies, in this work it is critical to identify how adaptations alter both kinetics and thermodynamics of activity and folding to reveal how mutations improve function and fitness.

---

[13] *Fundamentals of Enzyme Kinetics* by Athel Cornish-Bowden (1979) and *Comprehensive Enzyme Kinetics* by Vladimir Leskova (2003) are rich resources for the study of enzymes that have been helpful to me during my thesis work.

4.1 Transition state theory and temperature dependence on reaction rates

As stated in Chapter 3, enzymes affect the rate at which a chemical equilibrium is reached and not the equilibrium itself, which is defined by the free-energy change between reactants and products (ΔG), or in the case of enzymatic reactions, substrates (S) and products (P) (Figure 4.). Enzymes are catalysts because they lower the peak intermediate (transition state, ‡) energy between substrate and product relative to a non-catalyzed reaction. Lowering the transition state activation energy ($\Delta G^{\ddagger}$) required to break and form new chemical bonds, vastly increases the rate at which the equilibrium state is reached.

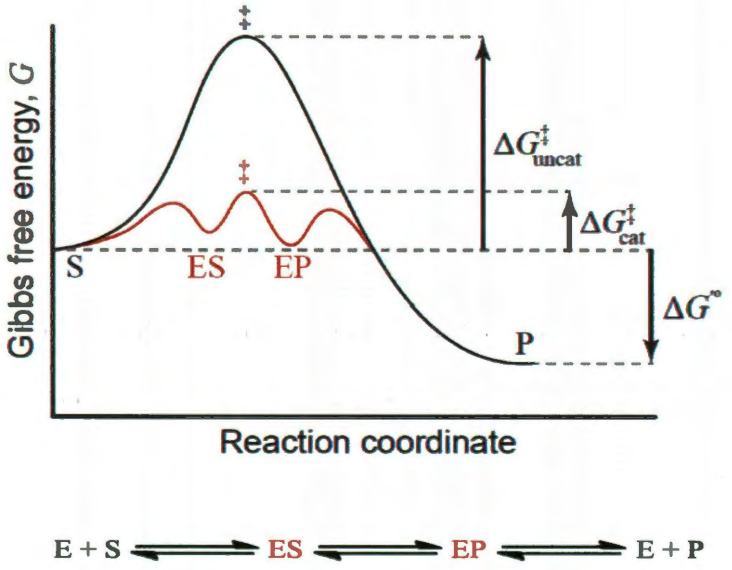


Figure 4.1
The smaller change in Gibbs free energy for catalyzed ($\Delta G^{\ddagger}_{cat}$) compared to uncatalyzed ($\Delta G^{\ddagger}_{uncat}$) reactions is responsible for increased rates according to transition-state theory. The equilibrium that is reached is determined by the difference in Gibbs free energy of the reactant and product ground states. Transition states on the reaction coordinate occur at peaks of Gibbs free energy (indicated by ‡) and intermediate states occur at local minima, designated here by ES and EP in the case of a reversible mechanism with two central complexes.

Transition state theory is used to describe the first-order rate constant (k) of a reaction in terms of the reaction's transition state activation energy, the temperature at which the reaction occurs (T), Boltzmann constant (k_B), Planck constant (h), and gas constant (R) (Equation 4-1).[14]

$$k = \left(\frac{k_B T}{h}\right) e^{-\Delta G^{\ddagger}/RT} \qquad 4\text{-}1$$

As the transition state activation energy decreases, as it does by the influence of an enzyme catalyst, the rate constant for the reaction increases, or in other words the reaction proceeds at a faster rate. From Equation 4-1, we can observe that reaction rates also increase with increasing temperature, the first of two temperature dependences of activity that will be described in this chapter. Reaction intermediates, such as those of enzyme catalyzed reactions (ES and EP), represent metastable minima on the reaction profile (Figure 4.1).

4.2 The Michaelis-Menten equation

Leonor Michaelis and Maud Menten (1913) are credited as "founders of modern enzymology" due to their rigorous studies on the kinetics of invertase (controlling for pH of reactions with acetate buffers) which demonstrated that initial reaction rates at various substrate and enzyme concentrations could be modeled by a simple kinetic law (Cornish-Bowden, 1979). Their proposed kinetic mechanism (Equation 4.2) is the basis of many more complex mechanisms that would later be described.

[14] The formulation provided by transition-state theory, derived primarily by Eyring (1935), describes an activation energy that differs from that seen in the Arrhenius equation ($k = Ae^{-Ea/RT}$), often used in similar analysis. In the Arrhenius equation, E_a corresponds to the standard enthalpy $\Delta H^{\ddagger}$ in the van't Hoff law whereas its analog is equal to $\Delta H^{\ddagger} + RT$ according to transition-state theory.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P \qquad 4\text{-}2$$

The above equation assumes that the first binding event of substrate (S) to enzyme (E) is reversible (E + S ⇄ ES). The second, catalytic reaction was assumed to be both slower (rate-limiting) and irreversible (ES → E + P). This is generally applicable to experimental conditions where little product has been generated, which is the case when determining initial rates, and thus equilibrium between the enzyme-substrate complex and product seen in Figure 4. (ES ⇄ E + P) is far from being reached. The enzyme exists in two forms: free (E) and in complex with substrate (ES). The enzyme, in the last step, is recycled and can participate in subsequent catalytic reactions. The Briggs-Haldane treatment (1925) does not assume, as Michaelis and Menten had done, that enzyme-substrate complex was in thermodynamic equilibrium with free enzyme and free substrate (following from $k_{cat} << k_{-1}$). Instead, Briggs and Haldane derived a more general formulation that removed constraints on the rate constants (k_1, k_{-1} and k_{cat}) and assumed that a steady state would be reached in which the enzyme-substrate complex concentration is constant (d[ES]/dt = 0) shortly after the initiation of the reaction. The form of the rate equation derived from the steady-state approximation is still referred to as the Michaelis-Menten equation (Equation 4-3).

$$v = \frac{k_{cat}[E_0][S]}{K_M + [S]} \qquad 4\text{-}3$$

In the Michaelis-Menten equation, the catalytic constant (k_{cat}) is the rate constant for the second, irreversible step and more generally defines the capacity of the enzyme-

substrate complex to yield product. The initial rate (v) is also a function of the substrate concentration ([S]), total enzyme concentration ([E_0]) and the Michaelis constant (K_M).

The fundamental properties of an enzyme are k_{cat} and K_M. The k_{cat} value is a first-order rate constant for product formation from the enzyme-substrate complex for the simple Michaelis-Menten mechanism. For more complex mechanisms, k_{cat} is a function of first-order and second-order rate constants, and thus cannot be attributed fully to any one rate within a kinetic step. More generally, k_{cat} describes the catalytic output under substrate saturated conditions. When a single catalytic step is rate-limiting, k_{cat} closely approximates the rate constant of this step. However, if a single catalytic step cannot be said to be predominantly rate-limiting, k_{cat} remains a measure of product forming capacity, although it is now a complex function of multiple rate constants. The product of k_{cat} and [E_0] is defined as V_{max} ($V_{max} = k_{cat}[E_0]$). This value, though not a fundamental kinetic property of an enzyme because of its concentration dependence, is used to describe the mathematical limit of the Michaelis-Menten equation as substrate concentration increases (Figure 4.2).

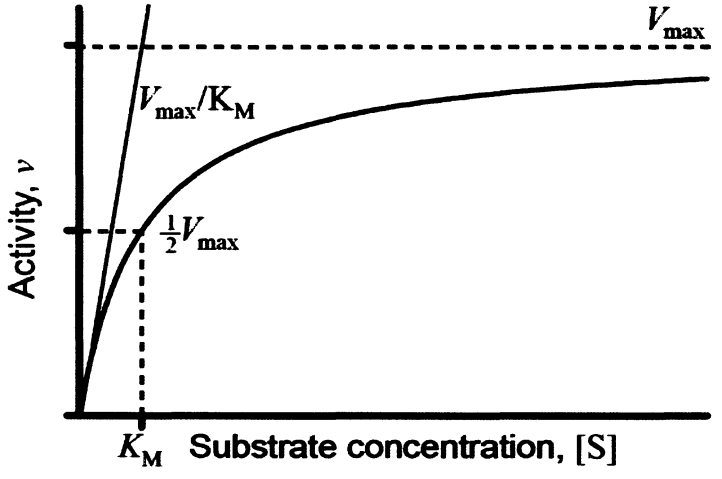


Figure 4.2
The Michaelis-Menten equation describes the response of enzyme activity measured as initial rates, v, to the concentration of substrate, [S]. Fundamental kinetic parameters (k_{cat} and K_M) are determined for a given enzyme and substrate by measuring initial rates at various substrate concentrations and fitting the Michaelis-Menten equation to the data. The curve defined by this equation is parameterized by the maximal rate, V_{max}, and the Michaelis constant, K_M. The value of V_{max} represents the limit of the Michaelis-Menten equation at increasing [S]. The substrate concentration required to reach half the maximal rate defines K_M for that substrate. The initial slope of the rectangular hyperbola, at low substrate concentrations ($K_M >> [S]$) is equivalent to V_{max}/K_M.

According to the Briggs-Haldane derivation, K_M is equivalent to the ratio of rates of decomposition of the enzyme-substrate complex to the rate of complex formation ($K_M = (k_{-1} + k_{cat}) / k_1$). Under the conditions Michaelis and Menten considered, where the catalytic step is rate limiting ($k_{cat} << k_{-1}$), the K_M can be approximated as the dissociation constant ($K_d = k_{-1} / k_1$). This result leads to the interpretation that the K_M is a measure of affinity, which is incorrect for any enzyme, possibly most enzymes, where the catalytic step is not rate limiting either by being much greater or comparable to rate of decomposition of the enzyme-substrate complex in the direction of the reactants. The K_M is determined by finding the concentration of substrate required for the initial rate to reach one-half V_{max}.

The specificity constant (k_{cat}/K_M), often described as a measure of catalytic efficiency or proficiency, represents the second-order rate constant at low substrate concentrations (Leskovac, 2003).[15] The physiological relevance of the specificity constant is that it is a measure of an enzyme's ability to discriminate between substrates in competition for the active site (Cornish-Bowden, 1979). It is for this reason that the specificity constant has been used to characterize promiscuous enzymes that undergo adaptation to more efficiently metabolize novel substrates (Copley, 2009). The specificity

[15] When $K_M >> [S]$, the Michaelis-Menten equation reduces to $v = (k_{cat}/K_M)[E_0][S]$.

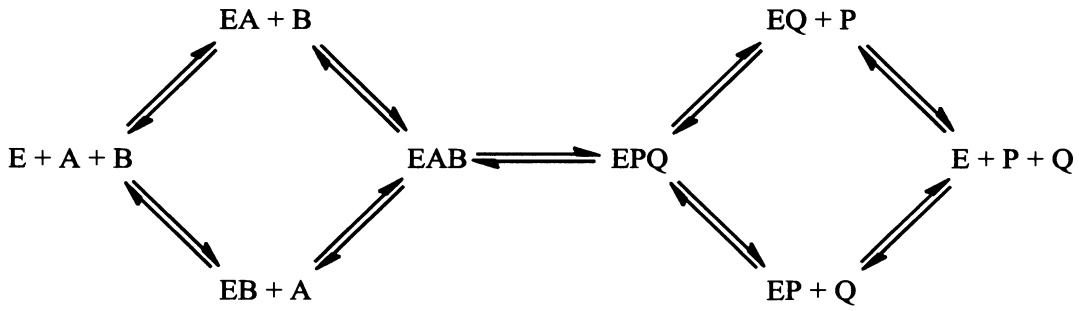
constant has been invoked as a proxy for fitness in cases where activity on a per molecule basis may be related to cell survival, which has been shown for some mechanisms of antibiotic resistance (Radika & Northrop, 1984; Soskine & Tawfik, 2010). However, a correlation between the specificity constant and fitness is yet to be firmly established, and it has been shown to lack significant correlation in various studies including those on antibiotic resistance (Brown *et al.*, 2010). Also, adenylate kinase can be said to act in saturating conditions (substrate concentration is milli-molar while K_M is on the order of micro-molar), which effectively decouples enzyme activity with substrate concentration by disrupting the linear correspondence. It is for these reasons that the specificity constant may have limited utility as a parameter used to differentiate AK variants.

It is important to note that the Michaelis-Menten and Briggs-Haldane formulations are incomplete (e.g. they do not account for fully reversible mechanisms with one or two central complexes).[16] Yet despite the simplicity of the formulation, the Michaelis-Menten equation has been extended to enzymatic reactions more complex than the two-step Michaelis-Menten mechanism used in its derivation. The equation captures the general rate-substrate curves (rectangular hyperbola seen in Figure 4.2) that characterize most enzymes, and it is therefore used as a first approximation for many reactions (Cornish-Bowden, 1979; Lehninger *et al.*, 2008).

[16] Athel Cornish-Bowden emphasizes in the 3rd edition of *Fundamentals of Enzyme Kinetics* (2004) that the Briggs-Haldane treatment is only a slight improvement on the Michealis-Menten formulation, and it suffers from many of the assumptions of the latter. All enzyme reactions are reversible, and yet both formulations have an irreversible product formation step that also display an asymmetric treatment of substrate/product bound states. A more accurate mechanism would be fully reversible with central complexes (E + S ⇄ ES ⇄ EP ⇄ E + P).

4.3 Application of Michaelis-Menten kinetics to the complex mechanisms of adenylate kinase

The reaction mechanism of adenylate kinase was elucidated by equilibrium isotope exchange and initial reaction rates studies (Rhoads & Lowenstein, 1968). It was shown that magnesium was required for activity, AMP and ADP could form inhibitory, non-productive ternary complexes, and that mechanism for adenylate kinase is random bi bi (Equation 4-4).[17]



4-4

The random bi bi nomenclature, established by Cleland (1963), describes a reaction pathway where the substrates (in this case AMP, ADP or ATP) may bind or release from the complex in a random, non-ordered sequence. In this mechanism, A and B refer to the reactants (AMP and ATP if assuming the forward reaction produces ADP) and P and Q refer to the products (two ADP molecules) of A and B, respectively. For a rapid-equilibrium random bi bi mechanism the reaction rate equation has been described:

[17] Inhibitory complexes have not been included in this mechanism for simplicity.

$$v = \frac{\frac{V_+[A][B]}{K_{iA}K_{mB}} - \frac{V_-[P][Q]}{K_{mP}K_{iQ}}}{1 + \frac{[A]}{K_{iA}} + \frac{[B]}{K_{iB}} + \frac{[P]}{K_{iP}} + \frac{[Q]}{K_{iQ}} + \frac{[A][B]}{K_{iA}K_{mB}} + \frac{[P][Q]}{K_{mP}K_{iQ}}} \qquad 4\text{-}5$$

where V_+ and V_- represent the maximal catalytic capacity of a given concentration of enzyme in the forward (+) and reverse (–) direction (analogous to V_{max} defined previously); [A] and [B] are the concentrations of the reactants and [P] and [Q] are the concentrations of products; K_{iA}, K_{iB}, K_{iP} and K_{iQ} refer to the dissociation constants of complexes EA, EB, EP and EQ respectively; while K_{mB} and K_{mP} are the dissociation constants for B or P respectively from ternary complexes EAB or EPQ seen in Equation 4-4 (Cornish-Bowden, 1979).[18] Under initial conditions, where little product has been formed, terms containing P and Q are negligible and can be omitted. Equation 4-5 further simplifies to Equation 4-6 by multiplying through by $K_{iA}K_{mB}$ (Cornish-Bowden, 1979).

$$v = \frac{V_{max}[A][B]}{K_{iA}K_{mB} + K_{mB}[A] + K_{mA}[B] + [A][B]} \qquad 4\text{-}6$$

If both reactants are at saturating concentrations, than the V_+ (in Equation 4-5) is equivalent to V_{max} used in single substrate equations. If one substrate is more abundant ($[B] >> [A]$), then terms lacking an expression for the concentration of said substrate can be deemed negligible, and if omitted we reproduce the Michaelis-Menten equation (4-3).

$$v = \frac{V_{max}[A]}{K_{mA} + [A]} \qquad 4\text{-}7$$

Here, K_{mA} is equivalent to the Michaelis constant for reactant A, when B is in excess. If the conditions had A in excess and B limiting, then Equation 4-7 could be

[18] For a random bi bi mechanism, it is assumed that dissociation constants for A and B (K_{mA} and K_{mB} or K_{iA} and K_{iB}), and similarly for the dissociation of P or Q, are interchangeable. In the simplification of Equation 4-5, $K_{iA}/K_{iB} = K_{mA}/K_{mB}$ follows from this assumption.

rewritten to express the reaction rate in terms of the Michaelis constant for substrate B. Thus, the Michaelis-Menten equation (4-3) can be applied generally to ternary mechanisms to determine the Michaelis constant for a reactant when the other reactant is in excess. I will use the Michaelis-Menten equation to describe AK activity at varying concentrations of ATP to determine k_{cat} and K_M for this substrate. However, at 60 °C and above it will become obvious that folding has a significant influence on activity. A novel mechanism that includes folding within the standard Michaelis-Menten kinetic scheme is required to determine the kinetic parameters of the mutants, and is one of the main objectives of my work, which will described in detail in Chapter 6.

4.4 Protein folding modeled as a two-state mechanism

Proteins are complex and dynamic three-dimensional structures whose folded conformations create unique chemical environments that catalyze specific reactions for particular substrates. From this it could be said that function follows fold. A properly folded state, which consists of a structural motif bringing together functional groups that form a reactive site, is the result of hydrogen-bonding, hydrophobic interactions, electrostatics, van der Waals interactions and opposing conformational entropy (Dill, 1990). Even interactions that make a small contribution to stability of the folded protein are significant due to the marginal stability of proteins at physiological temperatures (5-20 kcal/mol or protein) (Pace, 1975; Privalov, 1979). Even under conditions where a protein would be considered stable, some fraction of the ensemble of occupied conformational states includes those that are partially and fully unfolded.

Protein folding, like enzyme kinetics, is a complex phenomenon and scientists often apply simple thermodynamic and kinetic models to determine fundamental folding

parameter (ΔG, ΔH, ΔS, and ΔC_p) despite the underlying complex behaviors. Protein folding is often considered to be a spontaneous event, with the exception of cases where cofactors or covalent modification are required for maturation of the native state. Temperature, as well as pH, salts, pressure and chemicals denaturants, can alter the folded state of a protein in solution. The sharp, cooperative transition between folded and unfolded states observed by orthogonal methods (i.e. fluorescence quenching of aromatic amino acids, circular dichroism (CD), differential scanning calorimetry (DSC), etc.), supports a predominantly two-state mechanism for folding in some cases for simple proteins (Creighton, 1983). It is known that many small, monomeric proteins conform to a two-state mechanism (Jackson, 1998). This, of course, is an approximation, and even in the most well behaved cases one would not anticipate the thousands of bonds of the native protein to be formed simultaneously. There are many examples of proteins that do not follow an ideal two-state mechanism, and in general folding pathways are more complex (Jackson, 1998). Intermediates do exist along the folding pathway, however, they can be ignored and the two-state approximation applied if the transition states of unfolded conformations are short-lived relative to the transition state giving rise to the native state. The two-state folding mechanisms assume a single native state (N) and an unfolded state (U) that represents the remainder of the ensemble of conformations (the sum of all partially and fully unfolded states) (Creighton, 1988; Zwanzig, 1997).

$$N \underset{k_f}{\overset{k_u}{\rightleftharpoons}} U \qquad 4\text{-}8$$

The equilibrium constant between native and unfolded states (K_{unf}) is defined as the ratio of states or the ratio of the rates (k_u and k_f):

$$K_{unf} = \frac{[U]}{[N]} = \frac{k_u}{k_f} \qquad 4\text{-}9$$

The change in Gibbs free energy between unfolded and folded states can be described in terms of a folding reaction coordinate, which is analogous to the chemical reaction coordinate seen in Figure 4., and expressed as a function enthalpy (ΔH_{unf}) and entropy (ΔS_{unf}) or the unfolding equilibrium constant.[19]

$$\Delta G_{unf} = G_U - G_N = \Delta H_{unf} - T\Delta S_{unf} = -RT \ln K_{unf} \qquad 4\text{-}10$$

Zwanzig (1997) outlined assumptions required to justify the application of a two-state mechanism to the equilibrium unfolding of a protein: (1) there exists many unfolded states and a single native state, (2) the unfolded states exist in rapid equilibration and converge on a single, rate-limiting transition state culminating in a biologically active, native protein, and (3) no energetically trapped unfolded states persists nearly as long as the native transition state.[20] Assuming a two-state mechanism for folding transitions is rather convenient for determining thermodynamic parameters (K_{unf} and ΔG_{unf}).

Adenylate kinase has become a model enzyme for studying the interrelationships between catalysis, dynamics and folding (Henzler-Wildman & Kern, 2007). There are camps who choose to model the folding of AK as an apparent two-state mechanism for the purpose of comparing the thermodynamics and stabilities of mutants and orthologs (Bae & Phillips, 2004; Miller *et al.*, 2010). There are those primarily focused on identifying transient intermediates of AK along folding and dynamics pathways (Rhoades

[19] The equilibrium constant can also be defined as the inverse (K_f= [N]/[U]) which would result in the opposite sign of the respective change in Gibbs free energy (ΔG_{unf}= -ΔG_f= -(G_N - G_U)).

[20] The other assumptions describe the insensitivity of the folding rate constants to conditions and no one unfolded state is any more probable than any other. The latter of these is partially redundant with assumption number (3).

et al., 2003). And there are those interested in the confluence of both dynamics and catalysis (Wolf-Watz *et al.*, 2004; Henzler-Wildman *et al.*, 2007; Pisliakov *et al.*, 2009).[21] Due in part to the greater scrutiny of this model enzyme, it is known that AK does not strictly follow a two-state folding mechanism. For instance, Rhoades and colleagues showed using single molecule techniques that AK traverses the folding landscape from fully folded to fully unfolded and vice versa, not in a single step but by several intermediate steps, with some transitions taking greater than a second to complete (Rhoades *et al.*, 2003). What is the purpose of modeling AK folding as a two-state transition when it does not adhere to the assumptions implied in its application?

Although AK is not explicitly two-state, the influence of intermediates is often inconsequential to most thermodynamic studies, including this work. "All approaches to understanding the molecular basis of protein stability ultimately depend on reliable experimental determinations of the thermodynamics of protein unfolding for proteins of known structure" (Robertson & Murphy, 1997). Differential scanning calorimetry (DSC) can be used to determine the thermodynamics of protein unfolding, and is used in this thesis for this purpose. Fortunately, the calorimetric data analysis does not require an assumption to the presence of stable folding intermediates (Freire & Biltonen, 1978; Robertson & Murphy, 1997). There are many examples of proteins that do not have two-state folding kinetics (e.g. lysozyme which is said to fold with three-state kinetics) and deviate little from two-state calorimetric descriptions (Freire & Biltonen, 1978).

[21] I am referring to these as if they are distinct phenomena, when in fact they are two facets of the same.

B. subtilis AK does not strictly follow a two-state mechanism, and yet two-state models have been used in equilibrium unfolding and folding kinetic studies done on wild type and the isolated variants (Miller *et al.*, 2010). It may be that the transient intermediates of AK, including local conformational fluctuations, though identifiable by more sensitive methodologies (Rhoades *et al.*, 2003; Schrank *et al.*, 2009), may be negligible if the end states predominate. However, if the folding process is deemed to follow a non-two-state transition, DSC can be used to determine the number of intermediates and the thermodynamics of the intermediate states (Freire & Biltonen, 1978).

4.5 DSC analysis

Fitting the DSC data using a two-state non-zero ΔC_p model yields the thermal midpoint (T_m), enthalpy of unfolding (ΔH), and change in heat capacity for protein unfolding (ΔC_p).[22] The ΔH for a macromolecule is determined by integrating the excess heat capacity function. The shift in the baseline determines ΔC_p, which defines the temperature dependence of changes in enthalpy and entropy of unfolding.

[22] DSC data in this thesis is fit using a 2-state with non-zero ΔC_p using a Levenberg-Marquardt non-linear least-square method build in to the DSC Data analysis software in Origin. Further information regarding the fitting of the data can be found in the tutorial guide, version 7.0 (2004).

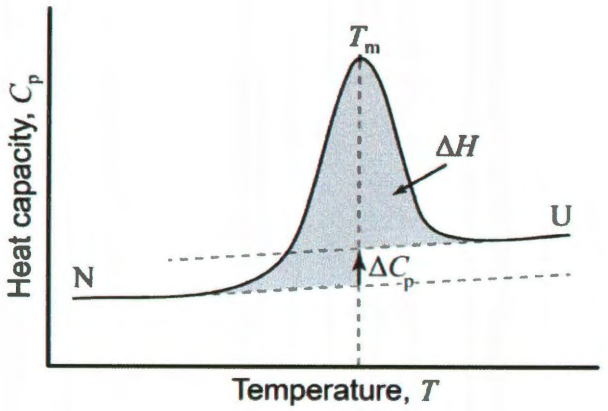


Figure 4.3
Analysis of differential scanning calorimetry data, consisting of measures of excess heat capacity as a function of temperature, can yield various thermodynamic parameters. The thermal transition of a protein from the native state to a fully unfolded state is characterized by its thermal midpoint (T_m), the change in enthalpy is given by the area under the heat capacity curve (ΔH), and the shift in baselines representing the native and unfolded states determines the change in heat capacity (ΔC_p).

It should be noted that the thermodynamic parameters calculated using a two-state fit are not used to extrapolate to values at much lower or higher temperatures, but are modeled near to the transition temperature where these values are more accurately determined. This effectively limits the propagation of deviations in ΔC_p into calculations of ΔH as ΔC_p is subject to large errors (Robertson & Murphy, 1997). The change in Gibbs free energy (ΔG_{unf}) is calculated following a modified Gibbs-Helmholtz equation (Becktel & Schellman, 1987):

$$\Delta G_{unf}(T) = \Delta H(T_m)\left(1-\frac{T}{T_m}\right) - \Delta C_p\left(T\ln\frac{T}{T_m} + (T_m - T)\right) \qquad 4\text{-}11$$

where $\Delta H(T_m)$ is the enthalpy change at the transition temperature. If ΔG_{unf} is negative, the unfolded state is favored. The value of ΔG_{unf} at T_m is invariably zero, meaning the equilibrium does not favor the native or unfolded state. The following equation describes

the relation between unfolding equilibrium constants (K_{unf}), ΔG_{unf}, folding rates, and the ratio of unfolded and folded enzyme (rewritten form of Equation 4-10):

$$K_{unf} = e^{-\Delta G_{unf}/RT} = \frac{[U]}{[N]} \qquad 4\text{-}12$$

Here we can see the second temperature dependence of enzyme activity. As temperature increases, the unfolding equilibrium constant reflects an increase in the fraction of enzyme in an unfolded state. The equilibrium constant is a function of temperature and ΔG_{unf}, which is also a function of temperature. Assuming the enzyme cannot perform catalysis in the unfolded state, the fraction of enzyme capable of carrying out catalysis decreases as a function of temperature. Thus, temperature dependent enzyme activity is the product of both catalytic and folding phenomenon.

4.6 Chapter summary

Michaelis-Menten kinetics can be used to parameterize enzyme activity (k_{cat} and K_M) and a two-state transition used in the analysis of DSC data to determine the thermodynamics of unfolding (primarily ΔG_{unf}). The complex enzyme kinetics of AK can be described by the Michaelis-Menten equation under experimental conditions where initial rates are measured when one substrate is limiting and the other is in excess. For studying the thermodynamics of AK, a two-state fit can be used to analyze the calorimetric data because known transient intermediates do not create significant deviations from the two-state behavior.

How do single adaptive mutations alter these kinetic (activity) and thermodynamic (stability) properties, and how are these properties correlated to improved fitness and presumably function? Isolating the properties or combinations of properties

that are most strongly correlated to fitness will be used to identify the molecular basis for these adaptations.

In Chapter 6, I will show how the concepts of enzyme kinetics and simple folding kinetics can be used together to better model function of enzymes at temperatures at which the majority of the enzyme ensemble is unfolded (above the thermal midpoint). This in itself is not unique. The unique contribution comes when this function is then tied to fitness, and we have the first instance of modeling evolution of variants in a population using both enzyme activity and folding.

Chapter 5 – Materials and Methods

5.1 Bacterial Strains and Media

5.1.1 *Escherichia coli*

Five *E. coli* strains were used for various applications in this work. XL1-Blue Supercompetent cells (Agilent) were used for general cloning purposes, specifically with transforming ligation products. DH5α chemically competent cells (Invitrogen) are suitable for high-quality plasmid preparations. One Shot® BL21 Star™ (DE3) (Agilent) cells are used for protein expression, the first step in the purification of adenylate kinase (see Section 5.3.1 AK expression in *Escherichia coli*).

E. coli cultures were grown in LB media (Milli-Q H_2O to desired volume, 1% tryptone, 0.5% yeast extract and 0.5% NaCl) with aeration by agitation at 37 °C. The recipe for solid LB includes the addition of 15 g/L Difco™ granulated agar (Becton, Dickinson and Company, Sparks, MD) before autoclaving. Hot, sterilized media with agar is agitated on an orbital shaker gently until the media cools to approximately 50-60 °C (hot to the touch, but not unbearably so), whereupon appropriate antibiotics are added, mixed, and poured into Petri dishes. Ampicillin and chloramphenicol were used at 100 and 15 μg/mL concentrations respectively in liquid and solid media.

5.1.2 *Geobacillus stearothermophilus*

G. stearothermophilus is a gram-positive, facultative thermophile and was the focus of early genetic studies of thermophily. *G. stearothermophilus* was chosen by Couñago *et al.* (2005) for a system in which to create a temperature sensitive

recombinant for primarily two reasons: the range of growth temperatures overlapped with the activity-temperature profile of AK_{BSUB}, and an efficient procedure for transformation had been previously established (Wu & Welker, 1989).

The transformation procedure was optimized by Wu and Welker (1989) for NUB3621, a strain derived from the parental strain NUB36, a soil isolate, mutagenized with MNNG (N-methyl-N'-nitro-N-nitrosoguanidine) (Chen *et al.*, 1986). The mutagenesis inactivated the endogenous restriction-modification system and created a spontaneous mutation conferring rifampicin resistance (Hsr^- Hsm^- Rif^r). In addition to the restriction-deficient phenotype, NUB3621 was chosen for protoplast transformation studies because the optimal regeneration temperature (60 °C) was closer to the transformation incubation and L colony growth temperature (50 °C) compared to the parental and sister strains (Wu & Welker, 1989)[23].

Growth of *G. stearothermophilus* cultures requires supplementation of media with a chelating agent and salt solutions (1.05 mM nitrilotriacetic acid, 0.59 mM $MgSO_4$, 0.91 mM $CaCl_2$ and 0.04 mM $FeSO_4$) (Chen *et al.*, 1986). Both TS medium (4% tryptone, 0.5% NaCl and pH adjusted with 10% w/v NaOH) and modified LB (1% tryptone, 0.5% yeast extract, 0.5% NaCl and pH adjusted with 10% w/v NaOH) are made using Milli-Q H_2O, pH adjusted with 10% w/v NaOH pre-sterilization and contain the strain specific supplements. Complete recipes can be found in the Appendix A.1. Appropriate antibiotics (7 μg/mL chloramphenicol, 5 μg/mL rifampicin or 5 μg/mL tetracycline) are then added to media prior to use for liquid media and immediately before warm media

[23] L colony here refers to colonies formed by L form cells which exhibit a deficiency in cell wall formation.

(55 °C) is poured into Petri dishes for solid media. Bacto™ agar (Becton, Dickinson and Company, Sparks, MD) is used as a solidifying agent for cultivation of *G. stearothermophilus*.

Cultures on solid media stored at 4 °C are kept no longer than a month for *E. coli* strains and a week for *G. stearothermophilus*. Recovery rates of viable cells from solid media decreases with time, especially so for *G. stearothermophilus*. All stock cultures are preserved in their respective growth medium containing 15% glycerol (volume by volume) at -80 °C.

5.2 Purification of DNA and sub-cloning methods

Plasmid DNA was harvested from *E. coli* and *G. stearothermophilus* using the Wizard® *Plus* SV Minipreps DNA purification system (Promega, Madison, WI). Genomic DNA from *G. stearothermophilus* is obtained using either a modified Genomic-tip 20/G kit protocol with QIAGEN (Halden, Germany) Genomic DNA buffers or a modified UltraClean® Microbial DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA).

Constructs used either for the expression of adenylate kinase or carrying a mutant *adk* library were created by sub-cloning DNA fragments amplified by the polymerase chain reaction (PCR). PCRs were performed with one of two thermal cyclers with most use coming from the latter: Primus 96 Plus thermal cycler (MWG-Biotech AG, Ebersberg, Germany) and Mastercycler proS (Eppendorf, Hamburg, Germany). Primers, templates, polymerases and thermal cycler conditions uses in various sub-cloning can be found in Appendix A.3. Denaturation and extension temperatures are specific to the DNA

polymerase used, while annealing temperature is in most cases optimized for yield of the amplified product.

All restriction enzymes and T4 DNA ligase used in sub-cloning were obtained from New England BioLabs (Ipswich, MA). Digested and amplified DNA products are resolved by gel electrophoresis using a 1% agarose matrix with ethidium bromide, 100 V potential and visualized using a gel documentation system under UV illumination. Ligation products were transformed into XL-1 Blue cells. All putative construct samples sequenced by Lone Star Labs, Inc. (Houston, TX) are prepared according to company specific recommendations.

5.3 Purification of adenylate kinase protein

5.3.1 AK expression in *Escherichia coli*

Transform One Shot® BL21 Star™ (DE3) chemically competent *E. coli* (Invitrogen, Carlsbad, CA) with a sample of pET vector containing the adenylate kinase gene of interest. A 50 mL volume of LB media containing appropriate antibiotics (e.g. ampicillin 100 μg/mL final concentrations for cells carrying a pET-11a expression vector) is inoculated with cells from a single colony harvested from the transformation. This culture is grown overnight at 37 °C and 225 rpm. Inoculate 2 L of pre-warmed LB containing antibiotic in a 4 L baffled Erlenmeyer flask with 25 mL of overnight culture. Save 1 mL for glycerol stock for future expressions.[24] The 2 L culture is incubated at 37 °C and 225 rpm. Add IPTG to a final concentration of 0.5 mM in the 2 L culture to initialize expression once the culture reaches an optical density (O.D.) of 0.6-0.8,

[24] An additional sample can be reserved for plasmid isolation and final sequencing of the gene being expressed.

measured as the absorbance at 600 nm. Spin down one O.D. of cells for each hour that passes after the addition of IPTG. These samples will serve to monitor hourly expression of AK by 12% SDS-PAGE. After 4 hours, the total volume of culture is spun down at 8000 g. Cell pellet may be stored at -80 °C indefinitely.

The cells harvested by centrifugation are resuspended in approximately 100 mL of ice cold Buffer A (50 mM Tris pH 7.5, 50 mM NaCl, 1 mM $MgCl_2$ 0.1 mM EDTA, and 0.3 mM DTT). The resuspension is maintained on ice and lysed by sonication using Heat Systems – Ultrasonic Inc. (Plainview, NY) Sonicator Ultrasonic processor Model W-375 with an ultrasonic converter rod model H-1A (Duty cycle 70%, output control 7, 5x30 s pulses). The lysate is clarified by centrifugation at 15,000 g at 4 °C for 1 hr. The resulting supernatant is carefully removed and diluted with Buffer A to a volume of 500 mL. The clarified lysate is first applied to a pre-equilibrated DE52 gravity column; the first of 4 chromatography steps.

5.3.2 Additional information chromatography and protein handling

During purification, all lysate and eluate samples are maintained on ice or at 4 °C. Purity of AK within a sample can be estimated by SDS-PAGE (12% acrylamide) stained with Coomassie Brilliant Blue solution. The yield of purified AK can be approximated by taking into account purity determined by SDS-PAGE and the concentration of AK calculated using the Beer-Lambert law and an extinction coefficient for AK (calculated value of 12345 $M^{-1}cm^{-1}$). However, these approximations improve as purity increases, and have little value during early purification steps.

An ÄKTA fast protein liquid chromatography (FPLC) unit and Unicorn^{TM} system control software were used for all FPLC work. It is imperative that all protein solution and buffers are filtered (0.22 μm) before they are used with an FPLC system. Between chromatographic purification steps, AK is dialyzed against an appropriate buffer at 4 °C using Spectra/Por 1 dialysis tubing (Spectrum Laboratories, Inc., Rancho Dominguez, CA) with a 6-8 kD MWCO.

5.3.3 DE52 chromatography

The DE52 column consists of pre-swollen microgranular diethylaminoethyl cellulose stationary phase that functions as an anion exchanger resin. *B. subtilis* AK has a theoretical isoelectric point of approximately 4.6, a calculated value dependent on model specific library of pKa values (ExPASy proteomics server). At neutral pH, AK is negatively charged, and binds the stationary phase. The preparation and use of a DE52 column is described below.

Weigh out 35 g of dry resin (Whatman International Ltd., Maidstone, England), and suspend in approximately 300 mL of 0.1 M Tris pH 7.5, 1 M NaCl buffter. Leave the suspension to sit for at least 15 min to allow the resin to settle, and then decant the supernatant. Repeat until the pH reaches 7.5, whereupon the supernatant is once more decanted and the resin is resuspended in 300 mL of 10 mM Tris pH 7.5, 50 mM NaCl, and 0.1 mM EDTA. Repeat once more with 100 mL of this same buffer. The resuspension will be poured into a glass column (e.g. Kimble/Kontes Flex-column 2.5x20) and let settle before use.

Equilibrate column with at least 5 column volumes of Buffer A. Load dilute cell extract onto column and collect flow-through. The column is then washed with 500 mL of Buffer A or until the OD_{280} decreases to a value less than 0.1 to ensure the removal of most unbound proteins. Elute using a Hoefer SG 500 gradient maker (San Francisco, CA) with equal volumes (100 mL) of Buffer A in the mixing chamber and Buffer B in the reservoir chamber to produce a linear gradient from 50 mM to 1 M NaCl in buffered solution. Singly charged, monoatomic chloride ion will outcompete a greater number of negatively charged proteins for the positively charged surface area of the resin as the concentration of NaCl increases. Fractions are collected in 5 mL aliquots and samples were then analyzed by SDS-PAGE gel to identify AK containing fractions.[25] These fractions are then pooled together and dialyzed over-night against 2 L of Buffer A containing freshly added 0.3 mM DTT at 4 °C. AK typically elutes between fractions 10 and 20.

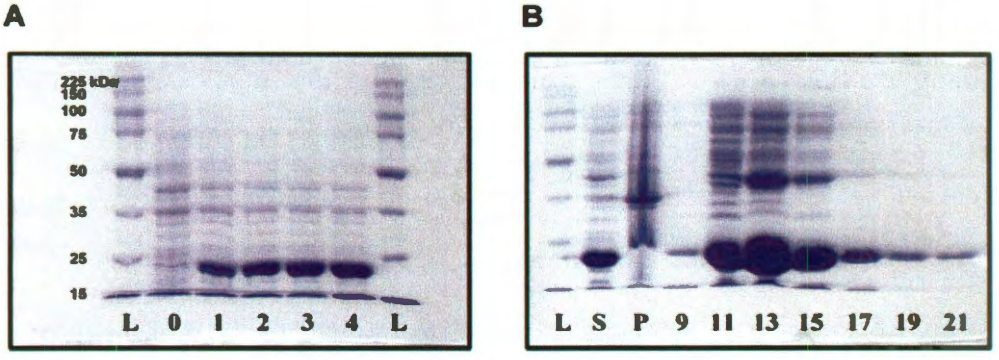


Figure 5.1
(A) Expression of *G. stearothermophilus* AK in One Shot® BL21 Star™ (DE3) cells is induced with the addition of IPTG (at time zero), and expression in samples are recorded

[25] One should exercise a discerning eye when choosing fractions to pool with the goal of avoiding contaminates. Choose purity over quantity as low yield is seldom an issue in purifying AK. The visible yellow hue associated with the AK containing fractions is most likely due to a co-purified contaminant as this color is absent after the final purification step, size exclusion chromatography.

hourly (0, 1, 2, 3 and 4 hours). (B) Proteins in lysate and eluate are separated by electrophoretic mobility by 12% acrylamide SDS-PAGE, and run alongside Perfect Protein Markers (designated L), 10-225 kDa (EMD Chemicals, San Diego, CA). Fractions 11-21 were pooled and dialyzed overnight.

5.3.4 Sepharose chromatography

The second stage of purification will utilizes a strong anion exchanger to further isolate adenylate kinase protein from other co-purified biomolecules. The dialyzed DE52 eluate is first filtered through a 0.22 μm filter (PES), as are all buffers used with a FPLC system. The HiTrap Q Sepharose XL column (GE Healthcare, Little Chalfont, England) is equilibrated with 10 CVs of degassed Milli-Q H_2O followed by 5 CVs of Buffer A using a FPLC system. If needed, two columns can be connected in series to increase binding capacity. The large volume of the eluate is then loaded onto the equilibrated column using a 50 mL Superloop (GE Healthcare) in conjunction with the FPLC. Non-bound protein was washed from the column with 5 CVs of Buffer A. The elution buffer transitions linearly from 100% Buffer A to 100% Buffer B over 10 CVs. The column is then cleaned with 2 CVs of 100% Buffer B. Elution of protein from the column is monitored using UV absorbance of eluate at 280 nm measured through a flow cell by the FPLC system. Fractions are collected in 3 mL aliquots, and AK purity of each ascertained by SDS-PAGE, pooled, and dialyzed against Buffer A as previously described. The majority of AK protein is often found between fractions 20 and 30, and fractions containing significant proportions of contaminants should be avoided.

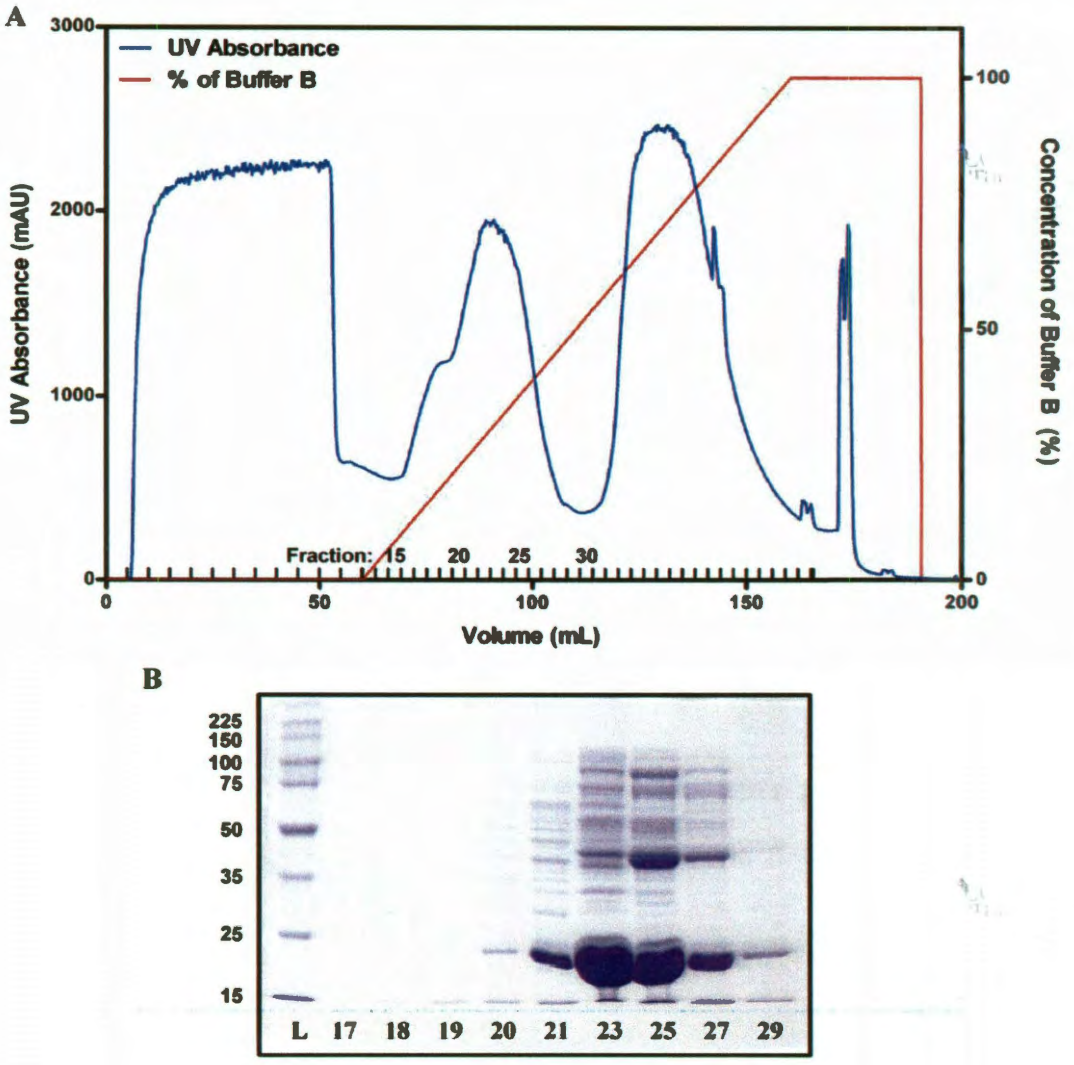


Figure 5.2

The second chromatographic step used to isolate AK utilizes a HiTrap Q Sepharose XL column in conjunction with an FPLC system. (A) The chromatogram records the absorbance at 280 nm of the eluate as a function of Buffer B, the elution buffer, and is simultaneously correlated to collected fractions. (B) Proteins in various fractions (17-21, 23, 25, 27 and 29) are separated by SDS-PAGE (12% acrylamide) to identify those most enriched in AK and lacking contaminants. Fractions 21-26 were pooled and dialyzed.

5.3.5 Affinity Chromatography

Affintiy chromatography is a common method for separating macromolecules by exploiting their unique affinities to ligands. Adenylate kinase from *E. coli* had been previously purified using a blue-sepharose column containing Cibacron dye (Barzu & Michelson, 1983). The Affi-Gel® Blue Gel employed here is a crosslinked agarose matrix with Cibacron Blue F3GA dye covalently attached (Bio-Rad Laboratories, Hercules, CA). The Affi-Gel® Blue Gel slurry of hydrated gel is poured into a 2.5x20 glass column and is then equilibrated with 500 mL of Milli-Q H_2O followed by 500 mL of Buffer A before applying sample. If the column has been used previously, elute tightly bound protein using the common buffer with a high ionic strength (2.5 M NaCl) and re-equilibrate the column. Once the sample has been loaded onto the column and flow-through collected, wash with at least two CVs of Buffer A. A 200 mL linear gradient from low (50 mM) to high (2.5 M) NaCl is used to elute protein from the column. Total protein concentration in various 5 mL fractions was determined by measuring the absorbance at 280 nm to aid in deciding which fractions to investigate by SDS-PAGE. Pool fractions and concentrate eluate to 2 mL using a Vivaspin 20 (10,000 MWCO PES) (Sartorius Stedim Biotech, Goettingen, Germany) centrifugal concentrator at 4,000 rpm using Beckman Coulter (Brea, CA) AllSpin JS-5.3 swing bucket rotor in an Avanit J-26 XPI centrifuge.

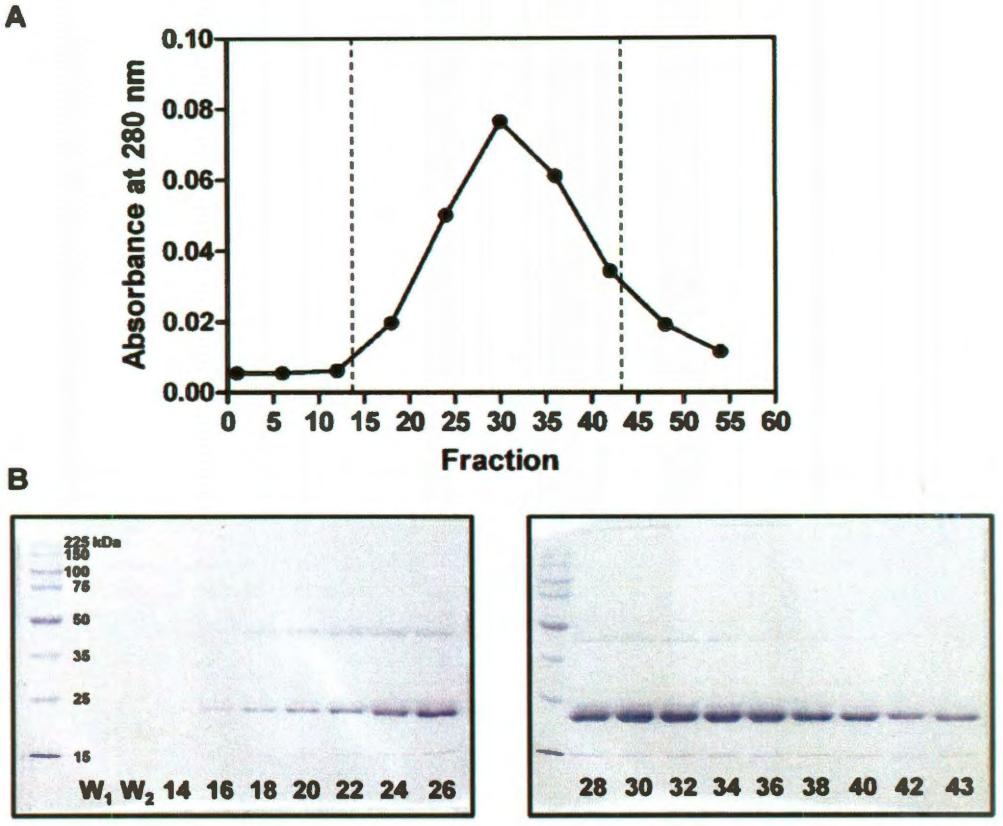


Figure 5.3
An Affi-Gel® Blue Gel column is used to separate proteins based on affinity for the third chromatographic step. Nucleotide binding proteins, such as AK, have a strong affinity to the Cibacron Blue F3GA dye in the stationary phase. (A) Absorbance at 280 nm is recorded manually for various fractions. (B) Wash through (W_1 and W_2) and even fractions about the absorbance peak are separated by SDS-PAGE (12% acrylamide) to identify those most enriched in AK. Fractions 28-45 were combined and dialyzed against Buffer A.

5.3.6 Size exclusion chromatography

Size exclusion, or gel-filtration, chromatography is a chromatographic method for separating macromolecules that have roughly twofold differences in size. Size dependent mobility complements the characteristics of charge and affinity that have been used in previous chromatographic steps to further purify AK from cellular extract. A HiLoad™

16/60 Superdex^{TM} 200 prep grade column (Amersham) is first washed with at least one column volume of filtered Milli-Q H_2O before equilibrating with Buffer A containing 150 mM NaCl. Filtered, concentrated protein solution is loaded, and buffer is run through the column at a rate 0.5 mL per minute. 2.5 mL fractions were collected once protein began to elute from the column. Samples were analyzed by SDS-PAGE. Fractions containing the purest AK samples were pooled and dialyzed against 2L of 20 mM HEPES pH 7.0 overnight at 4 °C Concentrate the protein using the Vivaspin concentrator to approximately 1 mM (approximately 20 mg/mL) and flash freeze samples in liquid nitrogen and stored at -80 °C. The protein purity was assayed using SDS-PAGE and ratio of absorbances at 280 and 260 nm.

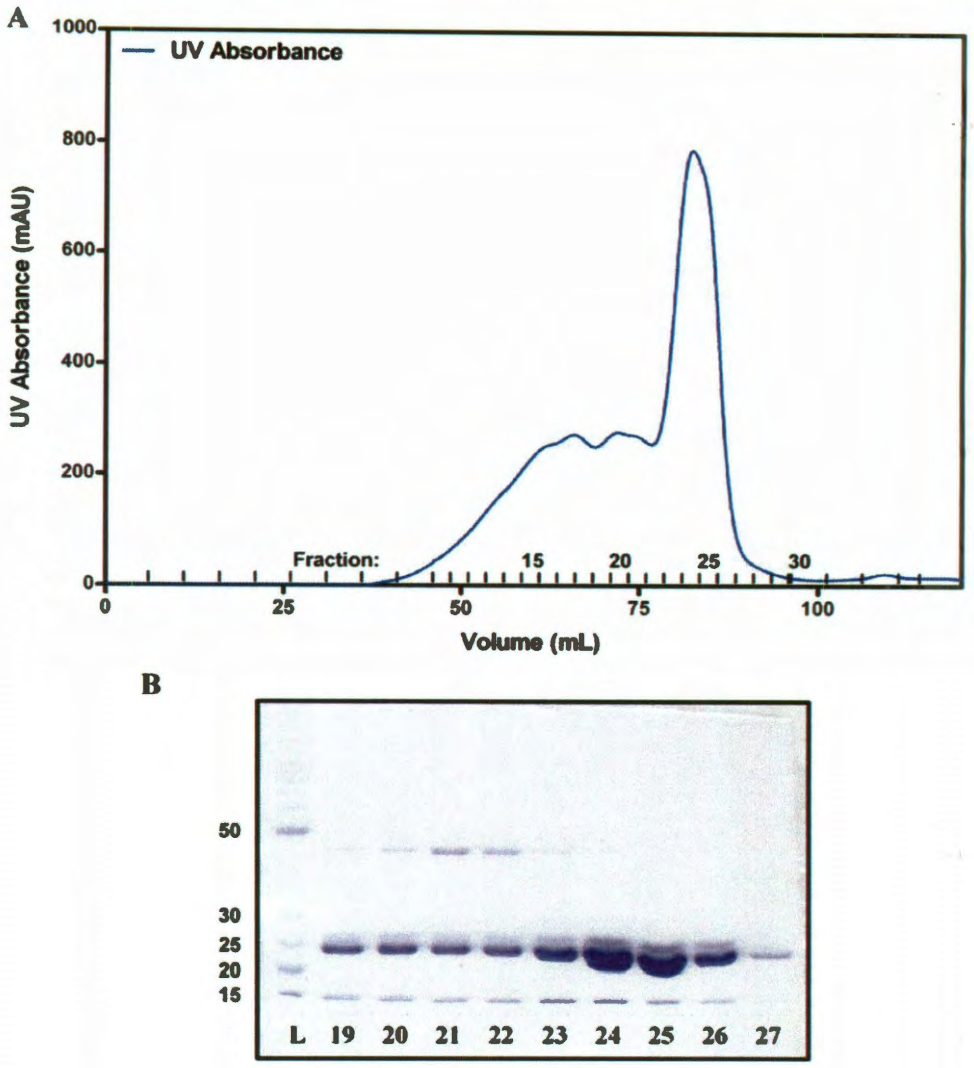


Figure 5.4

The last purification step isolates AK from contaminants by size exclusion chromatography controlled with an FPLC system. (A) The chromatogram records the absorbance at 280 nm of the eluate as a function of volume of buffer passed through the column. (B) Proteins in fractions 19 through 27 are separated by SDS-PAGE (12% acrylamide) to identify those most enriched in AK and lacking contaminants. BenchMark™ Protein Ladder (Invitrogen) is used as a molecular weight standard. Fractions 23-27 were pooled and dialyzed against 20 mM HEPES, pH 7.0.

5.4 Enzyme assay

Rates of ADP production for AK mutants were measured using an end-point coupled assay (Saint Girons *et al.*, 1987; Glaser *et al.*, 1992; Couñago *et al.*, 2008). The assay mixture contained buffer (25 mM phosphate buffer pH 7.2, 5 mM $MgCl_2$, 65 mM KCl), 1.4 mM AMP, and various ATP concentrations (5, 10, 25, 50, 100, 250, 500, and 1000 μM). The reaction was started by the addition of AK to a final concentration of 5 nM. The final reaction volume was 1.0 mL in a polyethylene tube. The reaction mixtures were incubated in a water bath at 55, 60, 62.5, 65, and 70 °C for five minutes prior to the addition of AK. The reactions were quenched 20, 40, and 60 seconds after initiation by rapidly adding 300 μl of the mixture to 100 μl of ice cold, 1.0 mM P^1, P^5-di(adenosine-5') pentaphosphate (Ap_5A).

The samples were maintained on ice until the amount of ADP could be measured. ADP was quantitated by a secondary assay in which 5 Units of pyruvate kinase, 0.3 mM β-nicotinamide adenine dinucleotide (NADH) and 0.5 mM phosphoenol pyruvate (PEP) were added to the quenched reaction. Absorbance at 340 nm was measured before and after the addition of 5 Units of lactate dehydrogenase. The conversion of NADH to NAD^+ was used as a measure of the ADP produced by AK at various temperatures, substrate concentrations, and reaction durations. The slope of [ADP] versus time for these three time points was used as the rate of product formation, v, at the middle time point. All measurements were performed in triplicate.

Plots of v versus [ATP] were fit to Equation 6-3 using Prism (GraphPad, La Jolla, CA) with t = 40 s for the exponential term describing irreversible denaturation. Calculated values for K_{unf} at each respective temperature for each mutant were used as

constants in the fit and k_{irrev} was shared across all temperatures for each individual mutant to constrain the fits. AMP, ADP, ATP, Ap_5A, NADH, PEP, pyruvate kinase, and lactate dehydrogenase were all purchased from Sigma-Aldrich (St. Louis, MO).

5.5 Differential scanning calorimetry

Measurements were performed using VP-DSC differential scanning microcalorimeter (Microcal, Northampton, MA) at a scan rate of 90 °C/hr. Data could be well modeled as an apparent two-state transition under the conditions of the DSC experiment. To test this approximation, a study of the concentration and scan rate dependent properties of wild type and all the mutants was performed (Supplementary Data). Protein concentration was 40 μM in 10 mM HEPES pH 7.0. Before measurements, sample and buffer were degassed for 10 min at room temperature. A pressure of 2 atm was kept in the cells throughout the heating cycles to prevent degassing. A background scan collected with the buffer in both cells was subtracted from each scan. Data were analyzed with Origin 7.0 software (MicroCal, LLC, Northampton, USA) to obtain the transition temperature (T_m), the calorimetric enthalpy of thermal unfolding (ΔH), the heat capacity change at the T_m (ΔC_p), and the Gibbs free energy change (ΔG_{unf}) at each temperature T.

5.6 Determination of growth rates

G. stearothermophilus strain NUB3621-R and recombinant strains expressing *B. subtilis adk* with Q199R or Q199R/Q16L were plated on modified LB (mLB) with rifampicin (5μg/ml) and chloramphenicol (7 μg/ml) overnight at 55 °C. Single colonies were used to inoculate 10 mL of mLB with rifampicin at 55 °C for approximately 3 hrs

with shaking at 225 rpm ($Abs_{600} \approx 0.5$). 50 µl of the 3 hr culture was used to inoculate 50 mL of mLB and incubated at the desired temperature with agitation (~200 rpm) using a New Brunswick Scientific Gyratory Water Bath Shaker, Model G76 (Edison, NJ). Growth was monitored by as changes in optical density. Growth curves were determined by fitting to an exponential function. All LB media were modified with the addition of 1.05 mM nitrilotriacetate, 0.59 mM $MgSO_4$, 0.91 mM $CaCl_2$, and 0.04 mM $FeSO_4$ (Chen et al., 2006).

5.7 Transformation of *G. stearothermophilus* NUB3621-R

The procedure for transforming *G. stearothermophilus* NUB3621-R is adapted from Wu and Welker (1989) who were the first to apply a protoplast fusion protocol developed by Chen *et al.* (1986) for the transformation of protoplasts with plasmid DNA. Transformation occurs in two steps: PEG-induced transformation of protoplasts with plasmid DNA and regeneration of the cell wall of the transformed protoplast. The term *spheroplast* describes the altered, round appearance of cells upon disruption of cell wall integrity, and *protoplast* lack a cell wall entirely. Disruption or removal of the cell wall by lysozyme treatment is thought to improve uptake of exogenous DNA.

Due to the difficulty of transforming *G. stearothermophilus*, success will depend on the fastidiousness and consistency with which the protocol followed. All liquid and solid media are pre-warmed to appropriate temperatures prior to use. If plates are stored at 4 °C, they will be equilibrated to room temperature for at least an hour before being placed at higher temperatures (50-60 °C). A GD120 series water bath (Grant) is used to incubate whole cells at and protoplasts at 50 °C during all phases of the transformation procedure except those during active growth, centrifugation or phenotypic expression.

To begin, *G. stearothermophilus* NUB3621-R is streaked on a TS-Rif plate from a glycerol stock that is immediately replaced at -80 °C. The plate is incubated at 60 °C for 12 h, upon which point an inoculate loop is used to collect a loopful of cells from a region of confluent growth. The bolus of cells is resuspended in 1 mL of TS, and 50 µl of the suspension is used to inoculate 10 mL of TS in a test tube. This initial liquid culture is maintained at 60 °C with shaking at 225 rpm in a Thermo Scientific Forma Orbital Shaker, model 420 (Waltham, MA) for two hours. Inoculate 50 mL of TS medium in a 250 mL baffled Erlenmeyer flask with 50 µl of the initial culture which should be at approximately 0.5 A_{600}. The 50 mL culture is grown at 60 °C and 225 rpm until the cell density reaches an A_{600} between 0.9 and 1.0. Cells are collected by centrifugation from 14 mL of culture in 15 ml, polypropylene centrifuge tube (Denville Scientific, Inc., Metuchen, NJ) at 4500 g for 5 min at room temperature. The cell pellet is resuspended in 1.4 mL of mP medium by vortexing.

At this point, the cell density of the sample can be determined by plating serial dilutions on mLB and incubating the plates overnight at 60 °C. Dilutions should be prepared in sterile 1.5 mL centrifuge tubes as follows: 10 µl of cell suspension into 990 µl of TS medium for a 100 fold dilution (10^{-2}) and 100 µl of cell suspension into 900 µl of medium for a 10 fold dilution (10^{-1}). Plating 100 µl of a 10^{-7} dilution of cell suspension should yield roughly 30 colonies, which translates to 3×10^9 cfu/mL for the undiluted suspension. A cell density of at least 2×10^9 cfu/mL is necessary but not sufficient for cultivating transformants.

Lysozyme is added to the undiluted cell suspension to a final concentration of 100 ng/ml, and the suspension is incubated for 10 min at 50 °C and 130 rpm to generate

protoplasts.[26] The protoplast-lysozyme mixture is diluted with 5 mL of mP medium, and protoplasts are collected by centrifugation at 3500 g for 5 minutes. The supernatant containing lysozyme is decanted, and the pellet of protoplasts is gently resuspended in 1.3 mL of mP medium.

The protoplasts should be held at 50 °C and are ready to be used. However, a set of controls can be performed at this point to diagnose causes of low transformation efficiency. To determine the percentage of cells converted to protoplasts, plate serial dilutions made in mP media onto mLB medium, which is selective against protoplasts. The plates are incubated at 60 °C overnight. Colonies represent cells not converted to protoplasts, and so a percentage can be determined by subtracting the ratio of non-protoplasts to total cell density from unity.[27] The efficiency of protoplast regeneration can be found by plating serial dilutions onto RA plates. Whole cells with cell walls intact will not replicate on RA. Plates are incubated at 50 °C for 12 h followed by 60 °C for an additional 12-24 h. Either low survival rates or low efficiency of protoplast conversion after lysozyme treatment can limit transformation efficiency, and can be diagnosed by comparing both controls.

The prepared protoplast solution is distributed into 1.5 mL microcentrifuge tubes in 100 µl aliquots and held at 50 °C. Aliquots can be stored at -80 °C; however, the freeze-thaw process will reduce transformation efficiency. At least 1 µg of plasmid DNA

---

[26] Growth phase at which the cells are harvested, lysozyme concentration and temperature are variables that I believe most significantly influence transformation efficiency. However, if transformation efficiency is low, then lysozyme and treatment duration should be varied. I refer to cells post-treatment as protoplasts because Wu and Welker (1989) referred to them as such, despite not being able to distinguish between protoplast and spheroplasts by microscopic or selective methods.

[27] This calculation assumes no cells perished during the lysozyme treatment and centrifugation steps.

is mixed into the protoplast solution, and the entire volume is transferred into 0.9 mL of 40% PEG-6000 in a 15 mL polypropylene tube and gently, but thoroughly, mixed. The tube is the placed in an incubated gyratory shaker at a 45° angle and mixed for two minutes at 50 °C and at 130 rpm. The transformation mixture is diluted with 2.5 mL of mP medium and mixed thoroughly by tipping end to end. The protoplasts are collected by centrifugation at 3500 g for 5 min, the supernatant is discarded and residual PEG-media mixture is cleaned from the walls of the tube. For phenotypic expression of the resistance markers carried on the transformed plasmid, the protoplasts are gently suspended in 0.1 mL of mP medium and incubated for one hour at 50 °C and 130 rpm. Protoplasts are diluted with 0.4 mL of mP medium and plated on non-selective and selective RA plates with appropriate antibiotics to discern survival rates and isolate transformants. These plates are incubated at 50 °C for 12 h and then 60 °C for 12 to 24 h. The shift to a higher temperature (60 °C) is required for regeneration of the cell wall and should be used when recovering reverted protoplasts, or "regenerants." Colonies formed by regenerants and protoplasts can be easily distinguished by colony morphology, in particular a distinct difference in size as can be seen in Figure 5.5.

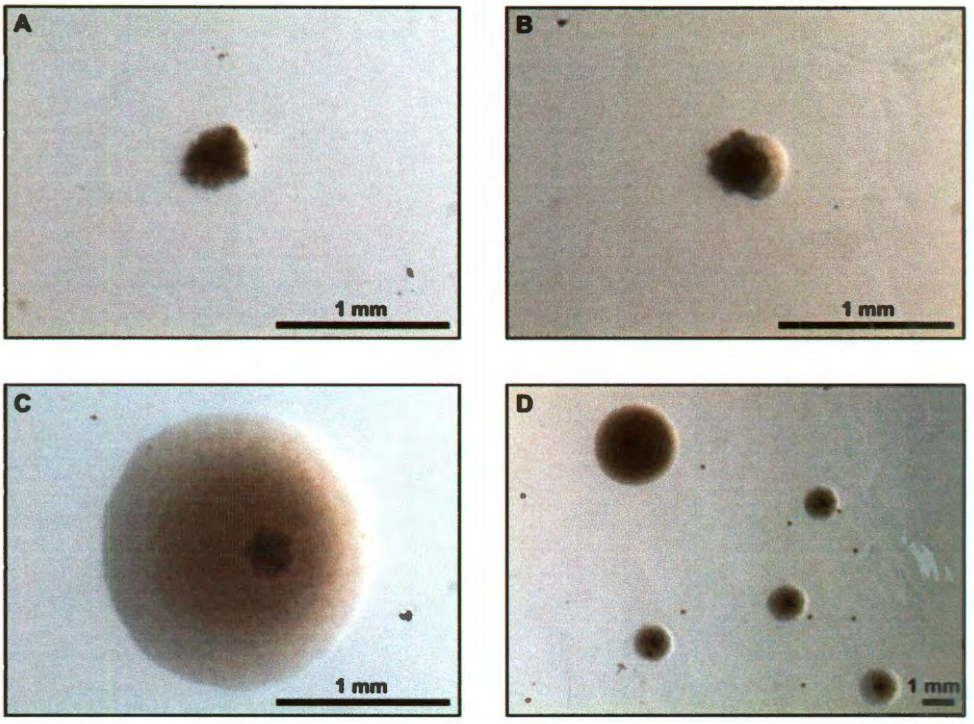


Figure 5.5
L-form and regenerant colonies are easily distinguished by size. (A) An L-form colony is opaque, punctiform and approximately 0.2 mm in size. (B) As protoplasts within the colony regenerate a peptidoglycan cell wall, cell proliferation increases and the colony exhibits two different morphologies. (C) A regenerants colony is characterized by a larger, translucent margin with a dark L-form remnant at its center. (D) A wide field view allows for contrasting sizes of L-form and regenerants colonies. All images are recorded after 36 h of incubation using an SZX16® research stereomicroscope (Olympus, Center Valley, PA) and captured with Go-3 digital camera and QCapture Pro 6 software (QImaging, Surrey, BC).

Chapter 6 – Relating evolutionary fates of mutants to *in vitro* activity

Much of the prose, figures and data contained in this chapter have been previously presented by Peña et al. (2010a) in the journal of Molecular Systems Biology (EMBO and Nature Publishing Group). The only portions reproduced here were those that were composed by the first author with the assistance of Yousif Shamoo. Quotation marks and references to this work have been omitted.

This work was done by myself, Milya Davlieva and Yousif Shamoo in collaboration with John S. Olson and Matthew R. Bennett, all of Rice University.

Couñago et al. (2005; 2006) used a 'weak link' method to favor mutations to an essential, but maladapted adenylate kinase gene within a microbial population. This experiment resulted in the identification of five mutants that arose nearly simultaneously and competed for success. To identify the biophysical basis for increased fitness resulting from adaptations to AK, it is necessary to marry ideas of enzyme catalysis and unfolding (described in Chapter 4) to provide and accurate description of activity as a function of temperature. A more complete description of activity that accounts for reversible unfolding and misfolding described herein can broadly be applied to studies of activity but is necessary for conditions where the enzyme under investigation favors the unfolded state.

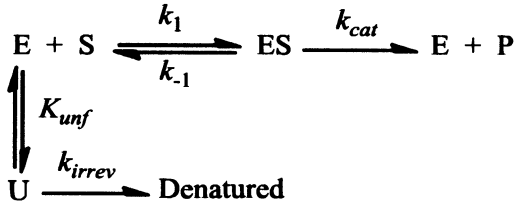
Physicochemical characterization of AK mutants showed that, although steady-state enzyme activity is important, success within the population is critically dependent on resistance to denaturation and aggregation. A fitness function based on *in vitro* measurements of enzyme activity, reversible and irreversible unfolding and the

physiological context reproduces *in vivo* evolutionary fates in the population linking organismal adaptation to its physical basis.

6.1 Irreversible folding pathway

The primary role of AK *in vivo* is to maintain adenylate homeostasis by catalyzing the reaction: ATP + AMP ⇄ 2ADP. We determined the catalytic activity of the adapted AK mutants using an *in vitro* end-point assay described in Section 5.4. The rate of ADP production was measured at saturating concentrations of AMP and varying concentrations of ATP (Saint Girons *et al.*, 1987; Couñago *et al.*, 2008). As described in Chapter 4, saturating concentrations of AMP simplifies the analysis so the Michaelis-Menten equation can be used to determine k_{cat} and K_M for ATP. To determine the dependence of catalytic rate on temperature, AK activity was measured at 55, 60, 62.5, 65, 67.5 and 70 °C for each mutant (Figure 6.1). At lower temperatures, the reaction rates show a hyperbolic Michaelis-Menten dependence on [ATP] for all the mutants examined; however, at higher temperatures, distinctly non-hyperbolic kinetics are observed (Figure 6.1). For example, the steady state kinetic pattern for AK_{BSUB} Q199R/A193V activity changes markedly as a function of temperature, showing a sigmoidal dependence on substrate concentration above 65 °C. At these elevated temperatures, the AK activity is close to zero at low [ATP] but is restored at higher substrate concentrations. Clearly ATP/AMP binding stabilizes the active form of the enzyme. Substrate stabilization, along with macromolecular crowding and chaperone activity, supports our observation that many of the AK mutants are viable *in vivo* at temperatures where the purified enzyme is largely denatured *in vitro* in the absence of ATP and AMP.

The distinct, sigmoidal shape of the reaction curves at higher temperatures indicates strongly that AK activity is lost by irreversible unfolding or aggregation in the absence of high concentrations of ATP. The observed sigmoidal, non-Michaelian kinetics can only be modeled assuming an irreversible reaction that directs active enzyme away from the catalytic pathway in competition with substrate binding (Equation 6-1, Scheme I) (Thomas & Scopes, 1998). The simplest mechanism that can explain the observed kinetics is shown in Scheme I below:



6-1

Scheme I

This mechanism utilizes two concepts discussed in Chapter 4 (Michaelis-Menten kinetics and two-state unfolding) and introduces a new one (irreversible denaturation or misfolding). This two-step irreversible denaturation process is known as the Lumry and Eyring model (Lumry & Eyring, 1954; Sanchez-Ruiz *et al.*, 1988; Sanchez-Ruiz, 1992). In this mechanism, the substrate-free enzyme in its active, native state, E, can reversibly and rapidly unfold into an inactive state, U. The equilibrium constant defining this unfolding reaction is K_{unf}, which is equal to [U]/[E]. This equilibrium ratio is unchanged throughout the reaction even though the absolute values of [U] and [E] are changing with both time and [S]. In effect, unfolding is competitively inhibiting substrate binding and can be overcome at high [S]. The net result is that increases in K_{unf} increase the apparent K_M for the reaction (i.e., $K_{M(obs)} = K_M(1+K_{unf})$). However, this effect alone cannot explain

the sigmoidal kinetics observed at high temperatures and low [S]. In our experiments, unfolded AK aggregates and becomes irreversibly denatured at a rate proportional to the fraction of unfolded enzyme, Y_U, during the assay.

$$Y_U = \frac{K_M K_{unf}}{K_M(1+K_{unf})+[S]} \qquad 6\text{-}2$$

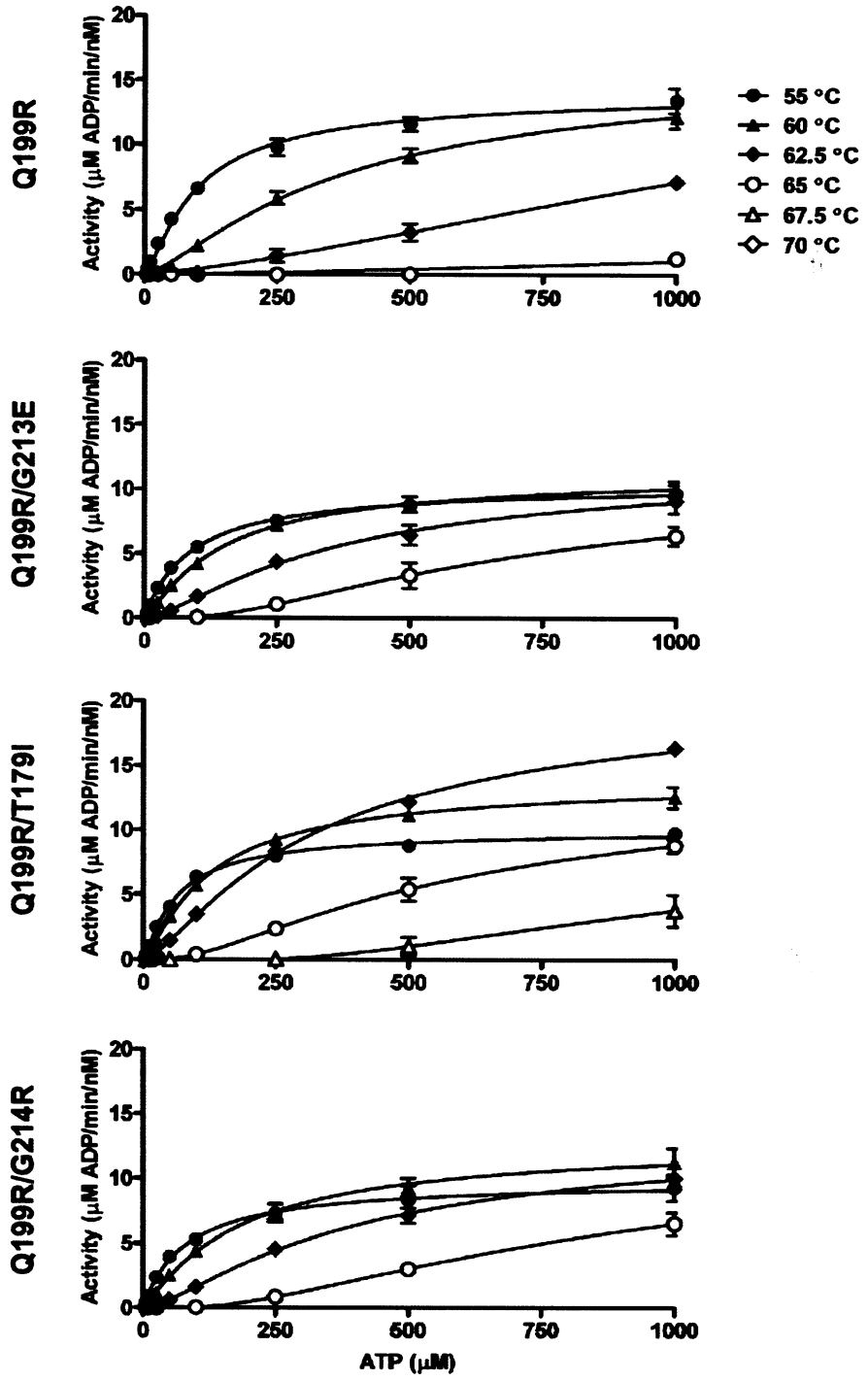
This *in vitro* process was modeled as an exponential decay of the unfolded state, i.e. $\exp(-k_{irrev} \cdot t \cdot Y_U)$ (see Appendix section A.4 for derivation and section A.5 for evidence of the exponential decay). The final initial velocity equation is:

$$v = \frac{k_{cat}[S][E]_0}{K_M(1+K_{unf})+[S]} e^{-k_{irrev} \cdot t \cdot Y_U} \qquad 6\text{-}3$$

The amounts of ADP formation at 20, 40 and 60 second time points were fitted to a straight line and the slope was used to determine initial steady state velocities, v. Plots of v versus [ATP] were fitted to Equation 6-3, allowing k_{cat}, K_M, and k_{irrev} to vary (Figure 6.1). The unfolding equilibrium constants at different temperatures were determined independently by differential scanning calorimetry (DSC). Because the AK reaction is a bisubstrate reaction with a ternary complex mechanism, the values of k_{cat} and K_M in Equation 6-3 depend on [AMP], which was fixed at 1,400 μM for all reactions.

As shown in Figure 6.1, Equation 6-3 predicts simple Michaelis-Menten kinetics at low temperatures when K_{unf} and Y_U are very small and $\exp(-k_{irrev} \cdot t \cdot Y_U) \approx 1$. However, when the temperature is elevated and K_{unf} becomes very large, the observed activity at low [ATP] is dramatically reduced by irreversible denaturation during the time interval between initiation of the reaction and the first measurement of product formation, ~40 seconds. However, at high [ATP] the fraction of unfolded protein is reduced markedly

due to ES complex formation and reduction in the fraction of the U state. As a result, irreversible denaturation is inhibited, enzymatic activity is retained, and the kinetics appear sigmoidal with increasing [S]. Substrate stabilization occurs through competition for free enzyme between the catalytic and unfolding pathways in a substrate concentration dependent manner. At high substrate concentrations, the catalytic pathway out competes the unfolding pathway. The fraction of unfolded protein, Y_U in Equation 6-3, is 0.94 for AK_{BSUB} Q199R in the absence of substrate at 55 °C, whereas in the presence of 1000 μM ATP, Y_U is only 0.02. This dramatic effect demonstrates the significant influence of substrate on the ensemble of folding states and the activity of the enzyme both *in vitro* and presumably *in vivo*.


Q199R
Q199R/G213E
Q199R/T179I
Q199R/G214R
Activity (µM ADP/min/nM)
ATP (µM)
0
5
10
15
20
250
500
750
1000
55 °C
60 °C
62.5 °C
65 °C
67.5 °C
70 °C

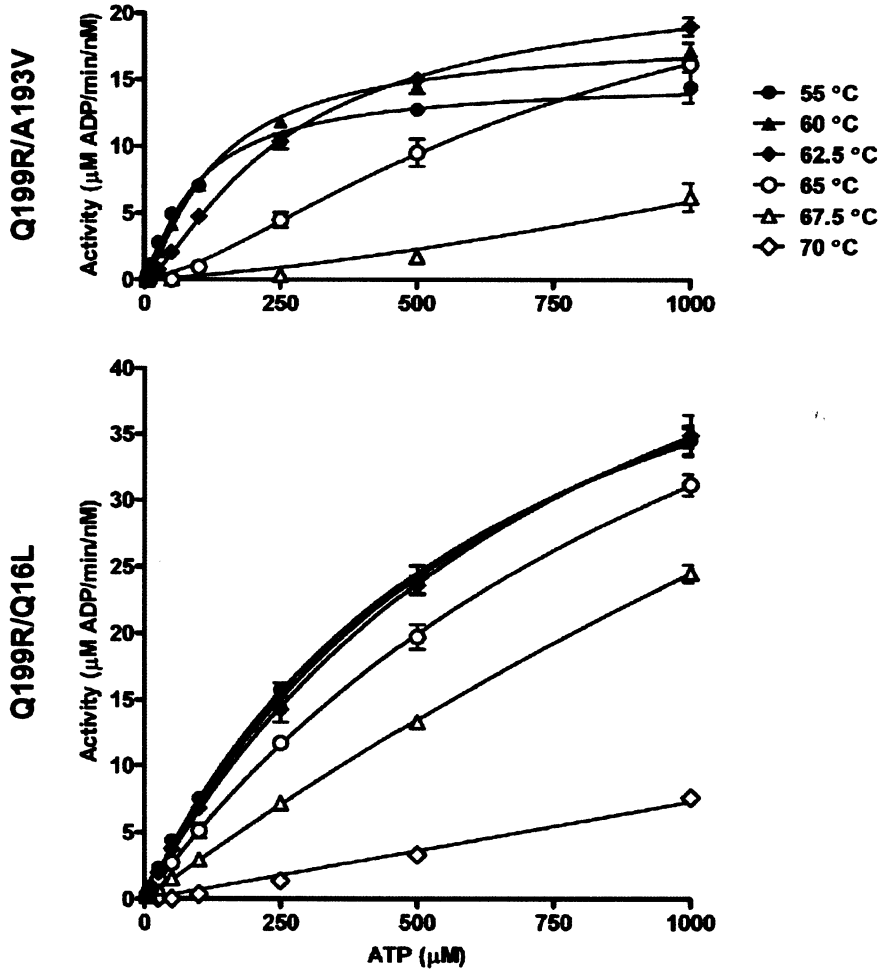


Figure 6.1
Enzyme activity of adaptive mutants (AK_{BSUB} Q199R, Q199R/G213E, Q199R/T179I, Q199R/G214R, Q199R/A193V and Q199R/Q16L) reveal a critical function for reversible and irreversible folding. The initial velocity is defined as the concentration of ADP produced in 1 min by 1 nm of AK. ADP production was determined at 5, 10, 25, 50, 100, 250, 500 and 1000 μM ATP with constant 1400 μM AMP. Symbols represent observed data points. Plots of initial velocity versus [ATP] were fit to Equation 6-3 and the fitted curves are presented as solid lines. The non-hyperbolic and sigmoidal curves observed at elevated temperatures are characteristic of an irreversible-unfolding pathway. Error bars represent the standard deviation of three measurements.

Table 6.1

Kinetic and folding parameters with standard deviations for AK_{BSUB} mutants at various temperatures determined by nonlinear fits of reaction rates as a function of ATP concentration*

Enzyme	T, °C	Activity k_{cat}, s^{-1}	K_M, μM	k_{cat}/K_M, $\mu M^{-1}s^{-1}$	v_i, [S]=1000 μM μM/min/nM AK	Folding K_{unf}‡	k_{irrev}§, s^{-1}
Q199R	55	14.2 ± 0.2	1.4 ± 0.3	9.9 ± 1.8	13.2 ± 1.0	16.6 ± 3.3	0.07 ± 0.01
	60	17.0 ± 0.7	0.9 ± 0.2	18.7 ± 4.1	12.2 ± 0.8	103 ± 34	
	62.5	29.9 ± 12.5	2.0 ± 1.1	14.6 ± 10.0	7.1 ± 0.3	269 ± 105	
	65	13.2 ± 224	2.0 ± 26.9	6.6 ± 140	1.3 ± 0.4	701 ± 324	
	67.5	†	†	†	†	1880 ± 1030	
	70	†	†	†	†	5120 ± 3150	
Q199R/G213E	55	10.0 ± 0.2	2.2 ± 0.4	4.5 ± 0.9	9.7 ± 0.7	8.1 ± 2.5	0.06 ± 0.01
	60	11.2 ± 0.3	1.0 ± 0.2	11.6 ± 2.4	10.0 ± 0.6	36.5 ± 21.5	
	62.5	12.9 ± 0.6	1.4 ± 0.3	9.3 ± 2.2	9.1 ± 0.9	79.9 ± 62.3	
	65	23.9 ± 9.2	3.1 ± 1.7	7.7 ± 5.1	6.4 ± 0.7	175 ± 159	
	67.5	†	†	†	†	392 ± 441	
	70	†	†	†	†	893 ± 1140	
Q199R/T179I	55	9.9 ± 0.2	1.8 ± 0.2	5.7 ± 0.8	9.7 ± 0.2	6.02 ± 1.20	0.09 ± 0.01
	60	13.7 ± 0.2	0.7 ± 0.1	20.5 ± 2.8	12.6 ± 0.9	32.4 ± 14.7	
	62.5	20.9 ± 0.6	0.8 ± 0.1	26.6 ± 4.0	16.4 ± 0.1	76.4 ± 43.5	
	65	17.1 ± 1.2	0.9 ± 0.2	19.3 ± 3.8	8.8 ± 0.5	181 ± 127	
	67.5	49.6 ± 198	2.5 ± 6.9	20.2 ± 98.2	3.8 ± 1.2	429 ± 355	
	70	†	†	†	†	1020 ± 987	
Q199R/G214R	55	9.5 ± 0.2	3.5 ± 0.7	2.7 ± 0.5	9.3 ± 1.0	4.57 ± 0.21	0.06 ± 0.01
	60	12.5 ± 0.3	1.3 ± 0.3	9.8 ± 2.0	11.2 ± 1.1	30.0 ± 3.6	
	62.5	14.5 ± 0.9	1.5 ± 0.4	9.7 ± 2.4	10.0 ± 0.5	77.5 ± 14.0	
	65	39.4 ± 39.9	4.1 ± 4.3	9.6 ± 13.9	6.5 ± 0.9	207 ± 55	
	67.5	†	†	†	†	527 ± 156	
	70	†	†	†	†	1390 ± 488	
Q199R/A193V	55	15.1 ± 0.3	8.2 ± 0.9	1.9 ± 0.2	14.4 ± 1.2	2.36 ± 0.29	0.08 ± 0.01
	60	18.6 ± 0.3	1.6 ± 0.2	11.3 ± 1.3	17.1 ± 0.6	17.1 ± 5.2	
	62.5	23.9 ± 0.5	1.3 ± 0.2	18.7 ± 2.4	19.0 ± 0.7	47.3 ± 19.1	
	65	33.7 ± 2.5	1.6 ± 0.3	21.3 ± 3.9	16.1 ± 0.4	132 ± 67	
	67.5	70.9 ± 229	3.1 ± 7.7	22.7 ± 92.0	6.2 ± 1.0	381 ± 236	
	70	†	†	†	†	1100 ± 788	
Q199R/Q16L	55	57.1 ± 1.3	561.9 ± 24.9	0.10 ± 0.01	34.5 ± 1.1	0.13 ± 0.06	0.01¤
	60	60.6 ± 1.5	372.5 ± 16.6	0.16 ± 0.01	34.7 ± 0.7	0.72 ± 0.10	
	62.5	65.0 ± 2.4	249.6 ± 16.2	0.26 ± 0.02	34.9 ± 1.6	1.74 ± 0.08	
	65	69.3 ± 2.7	169.1 ± 10.2	0.41 ± 0.03	31.2 ± 0.8	4.37 ± 0.91	
	67.5	128.0 ± 13.4	248.3 ± 30.7	0.52 ± 0.08	24.5 ± 0.6	11.0 ± 4.1	
	70	~	~	~	7.6 ± 0.4	28.4 ± 15.4	

* The kinetic parameters for *B. subtilis* adenylate kinase mutants were fit to triplicate data with a quasi-steady state approximation (Equation 6-3).

† The enzyme has no measureable activity at the specified condition.

‡ K_{unf} was calculated using parameters determined by DSC.

§ k_{irrev} is fit as one value shared across temperatures.

¤ A value of 0.01 was used as a constant in the fitting scheme.

~ Value could not be determined.

Fits of the observed data to Equation 6-3 are remarkably good (Figure 6.1), particularly considering that K_{unf} was determined independently by DSC, and allow determination of k_{cat}, K_M, and k_{irrev} for the AK mutants that appear during adaptation to high temperatures (Table 6.1). In the case of the AK_{BSUB} Q199R/Q16L mutant, the observed kinetics showed high activity that depends almost linearly on the concentration

of ATP, making it difficult to define the individual k_{cat}, K_M, and k_{irrev} values. The linear kinetics for this mutant indicate that the reaction is effectively bimolecular with $k_{cat} >> k_{-1}$, and, therefore, the enzyme is not readily saturated at accessible concentrations of ATP.

A continuous function for activity with respect to temperature for AK that incorporates both reversible and irreversible protein folding is given in Equation 6-4. A temperature dependent function for activity is favored over an interpolation between measured activities because it provides a model that can be tested against evolutionary outcomes under various temperature regimens and provides a continuum of values that can be referenced by a fitness function (Figure 6.2). The temperature dependence of the catalytic constants and unfolding equilibria are expressed as:

$$v(T) = \frac{k_{cat}(T)[E]_0[S]}{K_M(1 + K_{unf}(T)) + [S]} \cdot e^{-k_{irrev} \cdot t \cdot Y_U(T)} \qquad 6\text{-}4$$

Linear fits of Eyring plots (1/T vs. $\ln(k_{cat}/T)$ of the catalytic constants (k_{cat}) of each mutant were used to construct a temperature dependent function for catalytic constants:

$$k_{cat}(T) = \left(\frac{k_B T}{h}\right) \cdot e^{-\Delta G^{\ddagger}/RT} \qquad 6\text{-}5$$

where k_B is Boltzmann's constant, h is Plank's constant, R is the gas constant, and $\Delta G^{\ddagger}$ is the free energy of activation. The k_{cat} values at the highest temperature where activity is often barely detectable were not used in these analyses due to the large uncertainties associated with the fitted values.

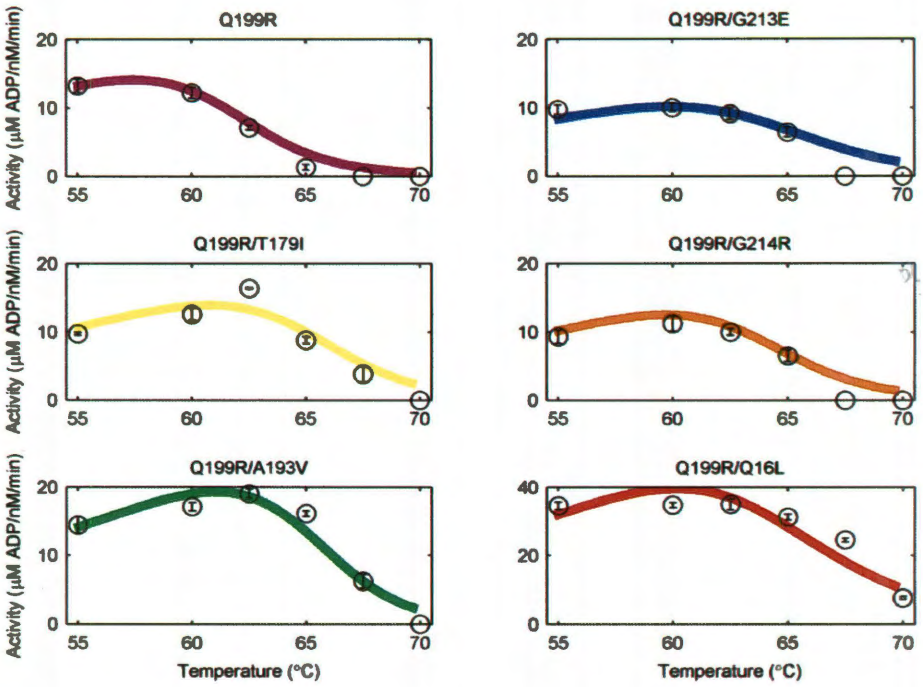


Figure 6.2
The temperature-dependent activity of the mutant AKs are very well reproduced by a continuous function that includes reversible and irreversible folding (Equation 6-4). The modeled reaction rates for various mutant AKs as a function of temperature are shown as solid lines and were used in the population simulation. Experimentally measured enzyme activities are represented by circles with standard deviations.

The calculated Gibbs free energy of unfolding (ΔG_{unf}) determined for temperatures between 55 and 70 °C were fit to a linear function in temperature. K_{unf} was calculated as (Equation 4-12 written in terms of T):

$$K_{unf}(T) = e^{-\Delta G_{unf}(T)/RT} \qquad 6\text{-}6$$

The temperature dependence of the fraction of unfolded enzyme (Y_U) was computed as:

$$Y_U(T) = \frac{K_M K_{unf}(T)}{K_M(1+K_{unf}(T))+[S]} \qquad 6\text{-}7$$

K_M is a complex term that can vary nonlinearly with temperature due to the temperature dependence of three constituent rate constants: k_1, k_{-1}, and k_{cat}. The measured K_Ms were relatively independent of temperature and thus averaged across experimental temperatures for a given mutant and used as a constant in the continuous temperature dependent activity function. The linearity seen by Arrhenius plots (1/T vs. $\ln(K_{M(obs)})$) in our previous studies can be accounted for by the temperature dependence of the unfolding equilibrium constant (K_{unf}) in the expression for the observed Michaelis constant ($K_{M(obs)} = K_M(1+K_{unf})$) (Couñago *et al.*, 2008). k_{irrev} is considered a constant over the temperature range examined.[28]

6.2 Thermal stabilities of mutant AKs

Experimental parameters obtained by DSC (midpoint, enthalpy change at T_m, and heat capacity change at T_m) were used to calculate the Gibbs free energy change for unfolding (ΔG_{unf}) at each temperature (Figure 6.3A and Table 6.2) (Becktel & Schellman, 1987). ΔG_{unf} was then used to calculate the unfolding equilibrium constants ($K_{unf} = \exp(-\Delta G_{unf}/RT)$) for use in analyzing the steady state kinetic data with Equation 6-3 (Table 6.1).

Table 6.2

Calculated Gibbs free energy change of unfolding, ΔG_{unf}(kcal/mol), at respective temperatures for each mutant with standard deviations.*

Enzyme	Temperature, °C					
	55	60	62.5	65	67.5	70
Q199R	-1.83 ± 0.13	-3.07 ± 0.22	-3.73 ± 0.26	-4.40 ± 0.31	-5.10 ± 0.37	-5.82 ± 0.42
Q199R/G213E	-1.36 ± 0.20	-2.38 ± 0.39	-2.92 ± 0.52	-3.47 ± 0.61	-4.04 ± 0.76	-4.63 ± 0.87
Q199R/T179I	-1.17 ± 0.13	-2.30 ± 0.30	-2.89 ± 0.38	-3.49 ± 0.47	-4.10 ± 0.56	-4.72 ± 0.66
Q199R/G214R	-0.99 ± 0.03	-2.25 ± 0.08	-2.90 ± 0.12	-3.58 ± 0.18	-4.24 ± 0.20	-4.93 ± 0.24
Q199R/A193V	-0.56 ± 0.08	-1.88 ± 0.20	-2.57 ± 0.27	-3.28 ± 0.34	-4.02 ± 0.42	-4.77 ± 0.49
Q199R/Q16L	1.35 ± 0.30	0.22 ± 0.09	-0.37 ± 0.03	-0.99 ± 0.14	-1.62 ± 0.25	-2.28 ± 0.37

* ΔG was calculated for each mutant at the listed temperature using Equation 4-11 and the parameters listed in Table A.2. The standard deviation is determined for three calculated values of ΔG based on three independent DSC measurements for each mutant.

[28] The irreversible rate (k_{irrev}) may be a temperature dependent phenomenon but was fit as a single value to limit the number of variables to be fit as a function of temperature.

AK unfolding can be well modeled as an apparent two-state transition, although re-folding is hampered by aggregation and partial loss of an intrinsic zinc ion (Bae & Phillips, 2004; Miller *et al.*, 2010). The thermodynamic parameters for each protein were determined as a function of concentration and scan rate to assess the extent to which AK and the associated mutants can be reasonably modeled as an apparent two-state transition (Appendix A.6). All of the isolated mutants have increased stability with shifts in T_m to higher temperatures relative to wild type AK (Figure 6.3A). As temperature is increased further, the instability or unfolding constants of the mutant proteins correlate well with the order in which their respective allelic frequencies decrease. The first mutant to appear, AK_{BSUB} Q199R, has a modest 2.8 °C increase in T_m compared to the original wild type enzyme (T_m = 44.0 °C). Subsequent double mutants, AK_{BSUB} Q199R/G213E, AK_{BSUB} Q199R/T179I, AK_{BSUB} Q199R/G214R, and AK_{BSUB} Q199R/A193V, provide 1 to 6 degrees of additional stabilization, with T_m values equal to 48.0, 49.5, 51.0, and 52.8 °C, respectively. The very successful AK_{BSUB} Q199R/Q16L mutant is dramatically more stable, with a T_m = 60.9 °C, which is 10 °C greater than that for its AK_{BSUB} Q199R progenitor.

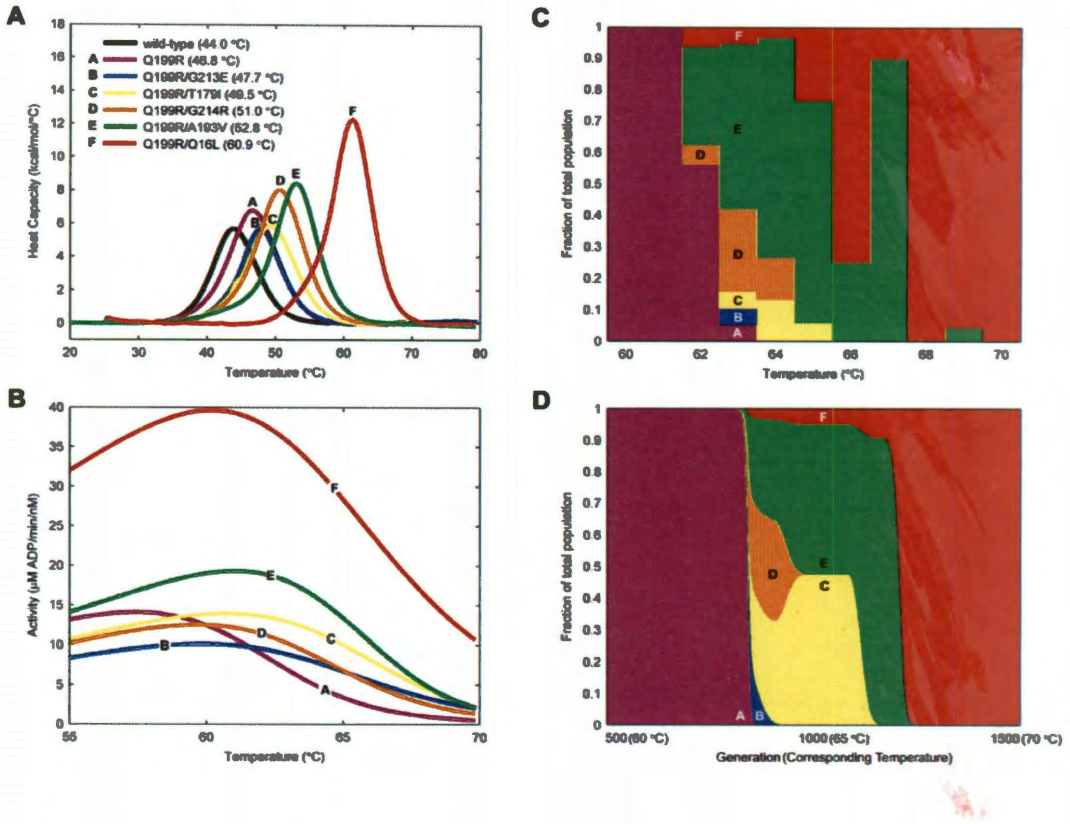


Figure 6.3

In vitro activity and stability can be used to accurately reproduce experimentally determined evolutionary outcomes during natural selection. (A) Stabilities of *B. subtilis* wild type and mutant AK as determined by DSC are presented here as the heat capacity as a function of temperature. The double mutants have midpoints or T_m values at higher temperatures compared to the AK_{BSUB} Q199R single mutant, which itself is ~3 °C more stable than wild type (T_m values are given in parentheses in panel A). The DSC data were used in determining K_{unf} for each mutant at various temperatures. (B) The kinetic (k_{cat} and K_M) and folding (k_{irrev} and K_{unf}) terms were used to fit ADP production rates for each mutant at 1000 μM ATP as a function of temperature. (C) The cumulative frequency of alleles obtained by experimental evolution within an exponentially growing population as a function of temperature. (D) The frequency of alleles determined from numerical modeling of the exponentially growing population using *in vitro* measurements. Above an activity threshold of 8.75 μM ADP/min/nM, relative increases in activity do not correspond to increases in growth rate and increased fitness. Below the threshold, there is a direct correspondence between growth rate and the AK activity presented in panel B. The colors used in panels A, B, C, and D correspond to the color scheme used to identify the positions of the mutations in Figure 1.

6.3 Stabilization of AK by ligand binding

Enzyme kinetics and calorimetric data fitted to Scheme I (Table 6.1) can be used to estimate the extent of stabilization that should be induced when 50 % of the protein is in complex with substrate. As a test of these estimates, DSC was used to experimentally measure the amount of stabilization induced by substrate binding (Figure 6.4). To preclude enzyme activity during the experiment, the transition state analog Ap_5A was used to measure substrate stabilization. DSC measurements at 0.25:1, 0.5:1 and 1:1 stoichiometric ratios of Ap_5A to AK_{BSUB} Q199R/A193V (Figure 6.4A) were performed. Under very tight binding conditions two populations with two melting temperatures corresponding to unbound protein (T_0) and the protein: Ap_5A complex (T_m) are observed (Figure 6.4A) (Brandts & Lin, 1990). Ap_5A affinity was estimated to be about 60 nM at 64.3 °C (Table 6.3). Ap_5A increased AK_{BSUB} Q199R/A193V stability by 12.3±1 °C. AK_{BSUB} Q199R/A193V showed little change in T_m as a function of Ap_5A concentration for the bound state, as is expected based on the theory explicated by Brandts and Lin (1990). Knowing the affinity of Ap_5A for AK and the T_m allowed us to compare the measured stabilization to that predicted from Equation A-7 which is in excellent agreement with the observed data (Table 6.3). Conversely, this data can be used predict the amount of Ap_5A that is required to shift the T_m of AK_{BSUB} Q199R/A193V to 64.3 °C. The predicted concentration of 12 μM Ap_5A is in excellent agreement with the experimental value of 10 μM.

A study of stabilization of AK_{BSUB} Q199R/A193V by ATP and AMP showed comparable results (Figure 6.4B), although these measurements should only be regarded as estimates due to enzyme activity (see Section A.7 for details). Again following

Scheme I, a solution of 20 μM AK_{BSUB} Q199R/A193V, incubated with either 1 mM ATP or 1 mM AMP would be expected to be stabilized by 8.0 and 2.4 °C respectively (Table 6.3). The extent of stabilization measured by DSC for AK_{BSUB} Q199R/A193V in the presence of 1mM ATP or 1 mM AMP was 7.1 and 1.4 °C, in good agreement with the expected extent of stabilization.

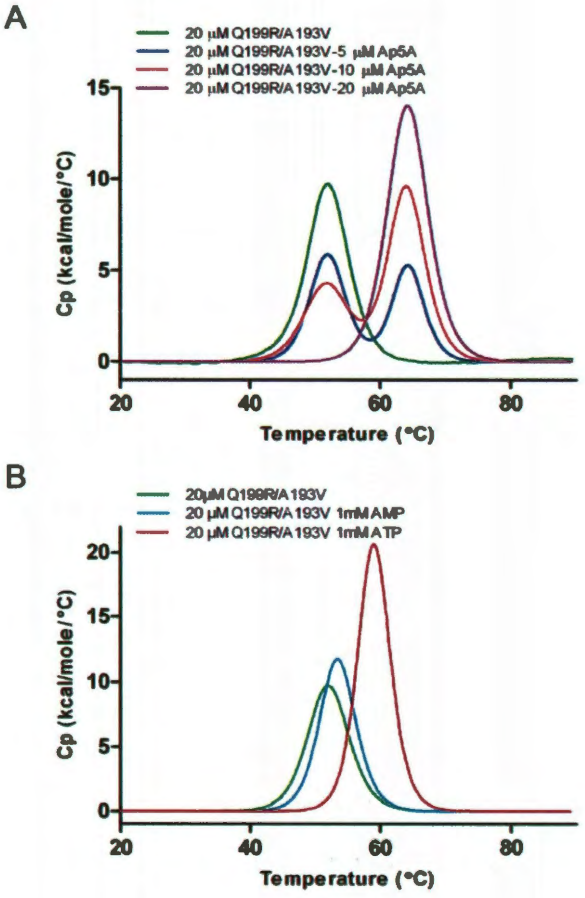


Figure 6.4
The extent of ligand induced stabilization monitored by DSC for AK_{BSUB} Q199R/A193V corresponds to values predicted by Scheme I (Equation A-7)). (A) Tight binding of the transition state analog Ap_5A shows distinct peaks with T_ms corresponding to the unbound and bound states in the expected stoichiometric ratios. (B) In the weaker binding case of nucleotides ATP and AMP, protein stabilization is observed as a single peak that has an increase in T_m that is dependent on ligand concentration (Brandts & Lin, 1990). Greater extent of stabilization is achieved with ATP relative to AMP due to a greater affinity for this substrate.

Table 6.3

Comparison of T_m values for AK_{BSUB} Q199R/A193V as determined by DSC in the presence of various ligands and those predicted by Scheme I and Equation A-7.

Ligand	K_B*, M^{-1}	[S]†, M	T_M‡, °C	
			Observed	Predicted
Protein alone	-	-	52.0	-
+ Ap_5A	1.6×10^7	1×10^{-5}	64.3	65.4
+ AMP	7.6×10^2	1×10^{-3}	53.4	54.4
+ ATP	1.7×10^4	1×10^{-3}	59.1	60.0

*K_B is the binding constant determined at 20 μM protein at a scan rate of 90 °C/hr.

†Estimation of the concentration of free ligand in the presence of 20 μM protein.

‡T_M is the thermal midpoint of a transition where approximately 50% of the ensemble is folded and bound in the presence of ligand.

6.4 Fitness function and population dynamics

The biochemical and structural properties of the AK_{BSUB} mutants provide the molecular basis for different physiological and metabolic phenotypes that are tightly linked to the fitness of the organism within the population. A single mutant and five double mutant alleles of *adk* were isolated during experimental evolution of a continuous microbial population under exponential growth conditions (Figure 6.3C) (Couñago *et al.*, 2006). Relative growth rates (fitnesses) are reflected by changes in frequencies of mutants within the polymorphic population over time. A fitness function consisting of temperature dependent kinetic and folding parameters was constructed and provides a biochemical description that can be used to describe quantitatively the experimentally observed population frequencies. This function also includes the particular properties of the turbidostat, provides a role for mutation rate in seeding the mutant subpopulations, and uses the product formation expression in Equation 6-4 as a fitness proxy for relative growth rates.

A numerical solution with discrete time steps was used to simulate the population dynamics of an exponentially growing polymorphic population as described in Appendix

A.8. The growth of each mutant subpopulation occurring at each generation time is given by:

$$P_i(t+1)=P_i(t)(1+r_i(t))+\sum_j m_{ji}r_j(t)P_j(t)-\sum_j m_{ij}r_i(t)P_i(t) \qquad 6\text{-}8$$

P_i is the frequency of a subpopulation i growing at a rate of r_i at generation t. The model includes terms whereby subpopulation i increases when mutations occur in the j populations forming more of the i population with a rate m_{ji} and decreases by drift away from i population by at a rate m_{ij}.

$$\text{If } v(T) \leq v_{threshold}, \text{ then } r_i(T) = \frac{v_i(T)}{v_{threshold}}.$$
$$\text{If } v_i(T) \geq v_{threshold}, \text{ then } r_i(T) = 1. \qquad 6\text{-}9$$

The relative growth rate (r_i) is defined as a value between 0, no growth, and 1, entire population replicates, and is the parameter that determines fitness selection. In our model, r_i is proportional to AK activity up to a specific threshold value, $v_{threshold}$, and is defined by $v(T)$ in Equation 6-3. Variation in the relative growth rates comes from the temperature dependences of K_{unf} and k_{cat}, which are measured independently *in vitro* (Table 6.1). The fraction of each subpopulation i is calculated as $f_i(t)= P_i(t)/P_{total}(t)$.

A fitness function based on enzyme activity and stability that includes an activity threshold, $v_{threshold}$, can accurately reproduce the frequencies and the order in which alleles are observed within the polymorphic, experimental population (Figure 6.3D). In this model, fitness is only proportional to $v(T)$ when activity falls below a critical physiological level (Equation 6-9). Increases in activity levels above the threshold have no corresponding increases in fitness and are not advantageous during competition

between subpopulations. If activity or stability were directly or largely proportional to fitness, AK_{BSUB} Q199R/Q16L would be the only double mutant observed at any of the elevated temperatures because this mutant is both highly active and very stable. However, significant transient subpopulations of AK_{BSUB} Q199R/T179I, Q199R/G213E, and Q199R/G214R do appear with increasing temperature, implying that only a threshold of activity is required for viability. Thus, the very favorable stability and kinetic attributes of AK_{BSUB} Q199R/Q16L serve to eliminate alternative models and help to define the critical physiological threshold for AK performance.

For polymorphism to persist within the rapidly growing population, the fitness differences between single nucleotide substitutions at the *adk* locus must be relatively small. Even the transient appearance of AK_{BSUB} Q199R/G213E at 63 °C corresponds to a substantial subpopulation of several billion cells that was reached within a total population of ~100 billion between the time it became more fit than AK_{BSUB} Q199R and less fit than the remaining double mutants. Under the condition where the relative fitness values of the mutants are very close, mutational frequency can influence the speed with which one strain is able to replace another. AK_{BSUB} Q199R/Q16L is one of the first three double mutants isolated at 62 °C but, despite its much more favorable kinetic and stability parameters, does not reach a majority status until 66 °C, which occurs 8 days or roughly 200 generations later (Figure 6.3C). AK_{BSUB} Q199R/Q16L is the only AK mutant that was the result of a transversion and suggests that AK_{BSUB} Q199R/Q16L could have been constrained by clonal interference if transitions occurred at a rate ten times greater than that for the transversion leading to AK_{BSUB} Q199R/Q16L. As shown in Equation 6-8, mutation rates are included within the general population model. The

prevalence of base transitions relative to transversions is well documented and has been observed for many organisms including *B. subtilis* and may certainly contribute to the observed evolutionary outcome (Wakeley, 1994, 1996; Sasaki & Kurusu, 2004).

6.5 Experimental tests of models: Substrate stabilization

Substrate stabilization of AK_{BSUB} contributes to the population dynamics observed in the numerical solution and presumably during natural selection. Increases in the concentration of ATP leads to prolonged persistence of Q199R, shifting the transition between allelic dominance to later generations in numerical models (see Appendix Section A.9 and Figure A.3). Marked increases in substrate concentration increase the activity of all the AK enzymes, eventually making them equally competitive when their activities surpass the threshold. The unfolding midpoints of almost all mutants in the absence of ATP/AMP (with the exception of Q199R/Q16L) are below the beginning experimental temperature. Q199R has a T_m of 47 °C, but the mutant strain was isolated in the culture up to 62 °C. Thus, substrate stabilization is almost certainly a major contributor to the activity of the mutant AKs at the temperatures above the T_m values of the apoproteins. Additional mechanisms of stabilization such as macromolecular crowding and chaperones may also be occurring. *In vivo*, a combination of these factors will produce additional stability, but in the *in vitro* assay increases in [S] clearly produce marked stabilization and retention of AK activity, which is reproduced quantitatively by our model.

Summaries of the stability and kinetic properties of the AK mutants are shown in Figure 6.3A and B, and the observed and calculated relative frequencies of each mutant population are shown Figure 6.3C and D. While there are many factors that determine the

effective concentration of intracellular proteins, the fitness function used to compute the theoretical populations in Figure 6.3D only required the relative *in vitro* stabilities and kinetic properties of the mutant enzymes to capture the general features of the *in vivo* evolutionary outcomes for the three major mutants AK_{BSUB} Q199R/A193V, AK_{BSUB} Q199R/T179I and AK_{BSUB} Q199R/Q16L and even for the limited transient success of AK_{BSUB} Q199R/G213E and AK_{BSUB} Q199R/G214R. Remarkably, at a given substrate concentration the only parameter that needed to be varied to accurately reproduce the experimentally observed population time courses was the activity threshold, $v_{threshold}$ (see Appendix A.10).

6.6 Experimental tests of models: Growth rates of AK mutants

Our physiological threshold model predicts equivalent growth rates between mutant strains at temperatures where adenylate kinase activities meet or surpass the threshold requirement. If the threshold is met, growth rates, i.e. fitness, will no longer be dependent on increases in activity of adenylate kinase, but on other limiting aspects of the metabolic pathways shared by NUB3621-R and the mutant strains, which differ only at the *adk* locus. Growth rates were determined for NUB3621-R along with mutants AK_{BSUB} Q199R and AK_{BSUB} Q199R/Q16L from 50 to 70 °C (Figure 6.5). Relative to *G. steraothermophilus* NUB3621-R, strains with AK_{BSUB} Q199R and AK_{BSUB} Q199R/Q16L have equivalent growth rates within the error of the experiment at low temperatures (50-55 °C) as predicted by the model. It is only at higher temperatures, between 55 and 57.5 °C, where AK_{BSUB} Q199R becomes less fit than the double mutant. AK_{BSUB} Q199R/Q16L mirrors the growth curve of NUB3621-R over the entire temperature range. This result is consistent with a recent experiment in which a starting

population of the mutant AK_{BSUB} Q199R/Q16L was grown at 50 °C for 10 days (~500 generations) to test whether new mutants would be fixed within the turbidostat at low temperature. An AK mutant with a 5 % or greater fraction of the total population would have been readily detected; however, none were observed (not shown), further confirming the predictions of the physiological threshold model and the growth rate study. A significant difference between growth rate of Q199R and Q199R/Q16L at 57.5 °C would suggest that the progenitor would be less fit before reaching 60 °C. In the original experiment of Couñago *et al.*, (2006) fixation of the double mutant strains is significantly slowed by clonal interference in the turbidostat population (Couñago *et al.*, 2006).

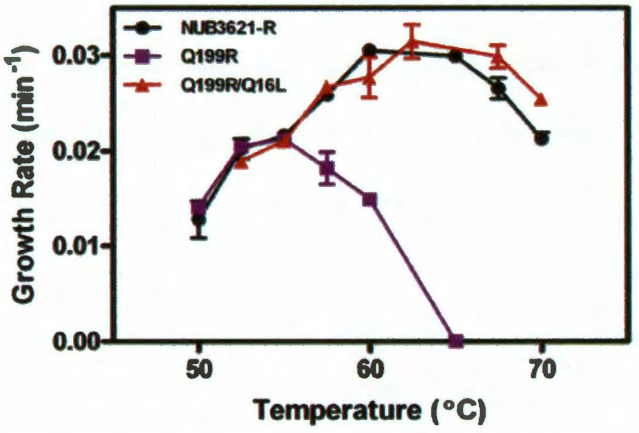


Figure 6.5
Equivalent growth rates at low temperatures support a physiological threshold for AK activity *in vivo*. Only at higher temperatures where the activity of AK_{BSUB} Q199R decreases below a threshold due to thermal denaturation do the growth rates, i.e. fitness, diverge from NUB3621-R.

6.7 Discussion

The successful adaptation of an organism during natural selection depends upon improvements in fitness that result from changes in molecular structure and function. In

our experiments, AK activity is tightly linked to the fitness of the organism and allows us to highlight the critical role of enzyme kinetics as well as reversible and irreversible protein folding to molecular adaptation. We have constructed a fitness function that incorporates classic Michaelis-Menten kinetics and protein folding properties to successfully reproduce the population dynamics of a microbial population during adaptation to higher temperature. As shown in Scheme I, the fraction of active, folded protein is dependent on both substrate stabilization and the probability that the protein becomes irreversibly lost from the system by aggregation or protein misfolding.

Misfolding and aggregation of proteins *in vivo* can have a substantial effect on fitness. As shown in Figure 6.1, it is essential to include a pathway for irreversible protein loss that reflects protein aggregation or misfolding as temperature increases. The irreversible unfolding step is analogous to the *in vivo* tendency of proteins to aggregate and can be an important factor in the evolutionary outcomes for a population with a particular mutant allele. The rate of this process, k_{irrev}, is required to appropriately account for enzyme loss in the *in vitro* assay and, when incorporated into the fitness function, represents the *in vivo* loss of physiological activity associated with aggregation or degradation. *In vivo* rates of irreversible unfolding will be affected by many factors including temperature, osmolytes and chaperones, each of which could play a role in adaptation (Tokuriki & Tawfik, 2009a). Even at lower temperatures, proteins sample unfolded or partially unfolded states and thus stability and denaturation kinetics are important and selectable properties. AK_{BSUB} Q199R/Q16L has the largest T_m among the selected mutants and is the only variant that is predominantly folded *in vitro* at 55 °C in the absence of substrate (~ 90 %). By comparison, less than 10 % of AK_{BSUB} Q199R

ensemble occupies the native state at 55 °C. *In vivo*, AK_{BSUB} Q199R is sufficiently folded and active to extend the viability of the organism by 9 °C and demonstrates a critical role for the *in vivo* folding environment.

An appealing aspect of our fitness function is that it permits an evaluation of specific and quantitative aspects of protein stability and activity to evolutionary fates. As shown in Figure 6.3, *in vivo* population dynamics can be reproduced using *in vitro* physicochemical measurements. Although *in vitro* measurements are made under relatively non-physiological buffer conditions, the relative activities and stabilities of the mutant AKs with respect to one another *in vivo* appears to be well captured as judged by the close correspondence of the observed and predicted population fractions in Figure 6.3C and D.

Our work demonstrates the importance of physiological context in defining fitness. Cellular energy charge (EC) is maintained by AK and is defined as: [ATP]+1/2[ADP]/[ATP+ADP+AMP] (Atkinson, 1968). When deficiencies in adenylate homeostasis limit growth rate, improvements in enzyme function can result in increased fitness but only to the extent that cellular needs are met, which is included as a physiological threshold in the fitness function for AK. In *G. stearothermophilus* the energy charge during exponential growth is 0.4-0.5 but falls dramatically and irrevocably once AK activity is lost (Couñago & Shamoo, 2005). AK alleles that are able to maintain EC do not limit cell growth because AK is no longer the 'weak link' in the cell. We believe this physiological threshold effect to be a general feature that defines and circumscribes the adaptive fitness landscape of protein evolution.

Adaptability is facilitated by increasing the diversity of the population and can be accomplished by the accumulation of near neutral or even modestly destabilizing mutations that provide more possibilities for success. Chaperones play an exceptionally important role in buffering biological systems against these destabilizing mutations as well as mistakes in translation that lead to polymorphic populations and have been shown to increase rates of adaptation (Rutherford, 2003; Drummond & Wilke, 2008; Tokuriki & Tawfik, 2009a). An increase in the diversity of the molecular ensemble permits accumulation of 'global suppressor' mutations that act largely on protein stability increasing the robustness of the molecule to subsequent adaptive changes (Bloom *et al.*, 2005; Bershtein *et al.*, 2008). Thus, adaptation through protein evolution is circumscribed by protein stability. As most mutational events will be destabilizing (Tokuriki & Tawfik, 2009b), higher mutation rates can lead to decreases in fitness eventually leading to extinction (Zeldovich *et al.*, 2007; Chen & Shakhnovich, 2009). Taken together, these studies and others suggest that the balance of molecular diversity and mutation rate is critical to understanding adaptation. While our system links the physicochemical properties of adaptive changes that increase stability, the principles apply equally to those changes that might decrease stability of the ensemble either through mutation or translational errors (Drummond & Wilke, 2008). Thus, regardless of how protein diversity is generated, evolutionary dynamics will likely be strongly coupled to stability and function.

Chapter 7 – Modeling energy metabolism through the phosphotransfer network

Much of the prose, figures and data contained in this chapter have been previously presented by Peña et al. (2010b) in the journal of Chaos (American Institute of Physics). The only portions reproduced here were those that were composed by the first author with assistance of Yousif Shamoo. Quotation marks and references to this work have been omitted.

This work was done by myself and Yousif Shamoo in collaboration with Elizabeth Van Itallie and Matthew R. Bennett, all of Rice University.

A central goal of our work has been the development of the 'weak link' approach that relates organismal fitness to the performance of a single gene. Using the 'weak link' approach, the scale of the problem is dramatically reduced but retains key aspects of the larger system. The intermediate 'steps' of adaptation can be easily identified and suitable models for adaptation developed within an experimental framework that promotes validation. In the 'weak link' approach, the one chromosomal copy of the essential adenylate kinase gene (*adk*) of the thermophile *Geobacillus stearothermophilus* was replaced with a copy from the mesophile *Bacillus subtilis* through homologous recombination. As *adk* is the only gene within the genome no longer adapted for growth at higher temperatures, the force of natural selection favors changes to *adk* as it is the 'weak link' within the genome of the organism. Creating the weak link effectively increases the flux control coefficient for the enzyme (introduced in Chapter 2), in this case AK. Therefore, adaptations to the weak link will return the greatest returns in flux through the affected pathway. Adenylate kinase maintains the crucial balance of ATP and ADP to AMP in cells and is therefore central to cellular energy homeostasis (as described

in Chapter 3). Cell death at non-permissive temperatures results from thermal unfolding of *B. subtilis* adenylate kinase (AK_{BSUB}) that, in turn, disrupts adenylate homeostasis.

During the course of selection the population was sampled and functional intermediates of adaptation observed as mutations to *B. subtilis adk*. The first mutant to reach fixation was a single mutation AK_{BSUB} Q199R. This allele was eventually replaced at 62-63 °C by five double mutants that arose nearly simultaneously within the population and share AK_{BSUB} Q199R as their progenitor (Figure 7.1A). All five double mutants arose to at least 5 % of the population in the temperature range of 62-63 °C as AK_{BSUB} Q199R declined (Couñago *et al.*, 2006). Reconstruction of the identified mutants by site-directed mutagenesis to the *B. subtilis adk* gene followed by homologous recombination within an unadapted strain recapitulated the phenotype (thermal tolerance) and reflects that these substitutions to *adk* and not potential mutations outside of this allele represent the predominant mechanism of adaptation to high temperatures. Adaptations to the growth environment could have occurred outside of *adk*; however, all adapted strains contained mutations to the *adk* allele indicating strong selection. These facts guided our efforts to quantitatively link fitness to enzyme activity; which in turn allowed us to reproduce evolutionary outcomes in a simulation that used activity (Figure 7.1B) as a proxy for fitness (Chapter 6). Changes to AK activity and thermal stability that resulted from mutation had direct consequences for cellular fitness and therefore met our goal for an experimental system that allows us to develop and test models for quantitative molecular evolution. The strong coupling between adaptations to AK and organismal thermal tolerance is due to the essential role of AK in energy metabolism. The fitness function developed to link activity to fitness was simple and effective, yet it could not

accurately capture the nonlinear response of fitness to enzyme activity because it did not relate AK activity to energy charge, a parameter which is a strong determinant for viability.

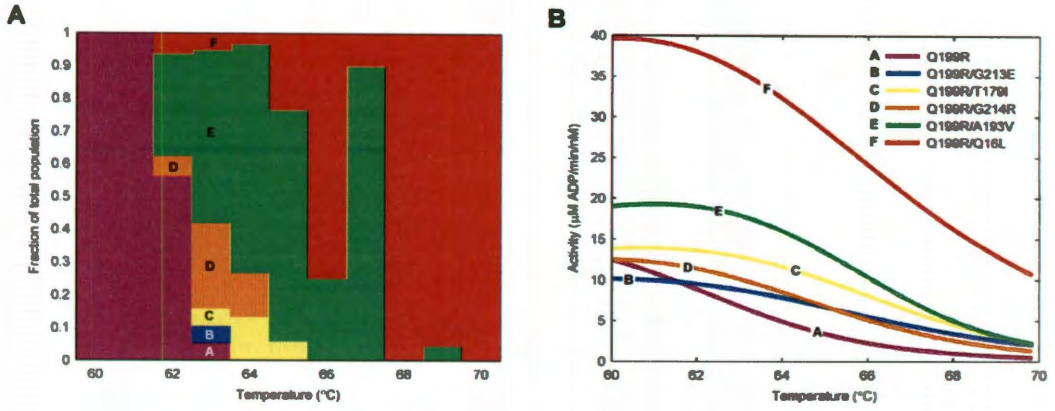


Figure 7.1
Fitness, responsible for changes in allelic frequency within a population is attributed to changes in enzyme activity. (A) The cumulative frequency of alleles obtained by experimental evolution within an exponentially growing population as a function of temperature (Couñago *et al.*, 2006). (B) AK activity and folding parameters were used to fit ADP production rates for enzyme variants at 1000 μM ATP as a function of temperature (Peña *et al.*, 2010a).

7.1 Model of adenylate kinase's role in adenylate homeostasis

Adenosine triphosphate (ATP) is the predominant energy carrier enabling cellular work to be directed through the hydrolysis of its two phosphoanhydride bonds. AK, in conjunction with ATP synthase, is responsible for ATP regeneration from lower energy moieties (ADP with a single high energy bond and AMP with none). AK catalyzes the reversible reaction that charges AMP with the addition of a phosphate forming an ADP at the expense of one phosphate bond from ATP:

$$AMP + ATP \rightleftharpoons ADP$$

ATP synthase utilizes the proton-motive force across the plasma membrane created by the electron-transport chain to drive the generation of ATP from the lower energy ADP:

$$ADP + P_i \Im ATP$$

Without AK, there is no efficient way of charging the AMP formed *de novo* or by reactions coupled to the energy releasing hydrolysis of ATP. High energy ATP is a fundamental regulatory effector of energy metabolism pathways (e.g. glycolysis, citric acid cycle and oxidative phosphorylation), and in some cases acts as an allosteric inhibitor slowing the rate of catabolic regeneration of ATP (e.g. phosphofructokinase within glycolysis). The relative intracellular concentrations of AMP, ADP and ATP reflect the overall energy level, and is often referred to as Energy charge (EC) (Atkinson, 1968):

$$\text{Energy charge} = \frac{[ATP]+1/2[ADP]}{[AMP]+[ADP]+[ATP]} \qquad 3\text{-}5$$

In the event that AK loses function, EC will decline due to a decrease in flux of ADP and consequently a decrease in ATP and a buildup of AMP. It is the essentiality of AK function that makes the weak link approach a successful experimental framework for the study of adaptation (Couñago & Shamoo, 2005; Couñago *et al.*, 2006; Couñago *et al.*, 2008).

7.2 Development of a fitness function linking AK activity to fitness

A combination of biophysical studies was used to evaluate the temperature dependent enzyme activity, folding and aggregation properties of the adaptive AK

mutants isolated during experimental evolution (Peña *et al.*, 2010a). In principle, the fitness of each cell carrying these adaptive changes should be largely dependent on changes to the physicochemical properties of the mutant proteins. A comparison of observed enzyme activity as a function of temperature and evolutionary outcomes supported the hypothesis that these enzyme variants have very comparable effects on fitness despite diverse molecular mechanisms for adaptation (Miller *et al.*, 2010). For example, AK_{BSUB} Q199R/Q16L, with the greatest *in vitro* activity at all measured temperatures, could not outcompete the other functional double mutants until they would later succumb to the effects of AK thermal denaturation (Figure 7.1). These observations suggested that enzyme activity alone is not a sufficient proxy for fitness. In the evolution experiment, a polymorphic population resulted across temperatures ranging from 62 to 69 °C over approximately 500 generations. Relative fitness differences are reflected in the frequencies of strains over time and temperature. Simulations have shown that fitness differences between variants must be small (less than 5% difference); otherwise this polymorphism would not have been so persistent.

To determine what parameters best correlate to fitness, fitness functions were constructed relating single parameters or combinations of parameters (e.g. k_{cat}, K_M and K_{unf}) to growth rates. These growth rates were then defined within an *in silico* population to determine which combination of parameters produced polymorphisms best matching empirical outcomes. Attempts to build a fitness function for these adaptive changes were only successful in reproducing the evolutionary outcomes when a 'physiological threshold' was applied to a model of *in vitro* enzyme activity (Figure 7.2) (Peña *et al.*, 2010a). The hypothesis was that once the physiological demand for adenylate

interconversion was met; further enhancements to AK function conferred diminishingly small returns on fitness. As the activity dropped below the threshold there would be a proportional and direct relationship between growth rate and activity.

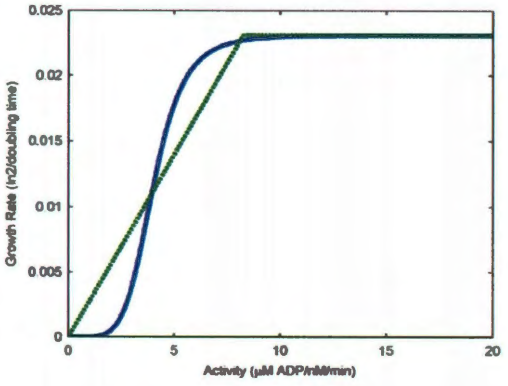


Figure 7.2
First approximation fitness functions include a simple physiological threshold (dashed) and Hill function (solid). A relative growth rate (y axis) can be calculated using these relations with an input of *in vitro* activity (x axis). Either can be used to reproduce experimental outcomes, but the Hill function has a form more consistent with response of the energy metabolism network to varying enzyme activity.

Although a 'physiological threshold' is simplistic, it was sufficient to reproduce the evolutionary outcomes in simulation (Peña *et al.*, 2010a). Such a 'physiological threshold', while useful, does not capture the central role of AK in the flux of adenylate species within the cell. For example, a loss of AK function below a certain point does not lead to relatively linear decreases in fitness, but rather results in a catastrophic and irreversible event - cell death. Simply, the organism is nonviable below a nonzero value of EC, and hence AK activity. To provide a more robust theoretical framework for future experimentation, we have constructed a fitness function modeling the response of energy charge at equilibrium to activity of AK.

A Hill function has qualities similar to our original fitness function with a threshold: increases in activity translate to negligible increases in growth rate far from the transition, and a sharp transition captures the physiological threshold required in our first approach (Figure 7.2). Hyperbolic or 'concave' fitness functions have been used previously to relate enzyme activity to fitness (Hartl *et al.*, 1985; Dean, 1995; Lunzer *et al.*, 2005) (See Chapter 2). These functions, equivalent in form to the familiar Michaelis-Menten equation, represent a special case of the Hill function (Equation 7-1) with a Hill coefficient of unity. A Hill function with a coefficient greater than one was needed to recreate the sustained polymorphism *in silico*, although attempts to match empirical frequencies resulted in a shift in the transition from AK_{BSUB} Q199R to double mutants to higher temperatures (64-65 °C) relative to the experimental shift (62-63 °C). To relate activity-temperature profiles of enzyme variants to competition between the mutant strains within a population, activities were fed into a Hill function to calculate a growth rate that was then used to determine population dynamics (Figure 7.3A). A Hill function with a midpoint (m) of 4.5 and a Hill coefficient of 5 (h) was used as the fitness function relating activity (a) to growth rate (w) reproduces evolutionary outcomes in Figure 7.3B consistent with experimental data seen in Figure 7.1A.

$$w = w_{max} \frac{\left(\frac{a}{m}\right)^h}{1+\left(\frac{a}{m}\right)^h} \qquad 7\text{-}1$$

This fitness function has been determined for one data set but can be used to predict outcomes of competition experiments between these adaptive mutant AKs or novel ones that will be characterized prior to competition. Despite the simplicity of the fitness

function, contrasted with the complexity of energy metabolism, we are able to reproduce evolutionary outcomes using only *in vitro* activity data.

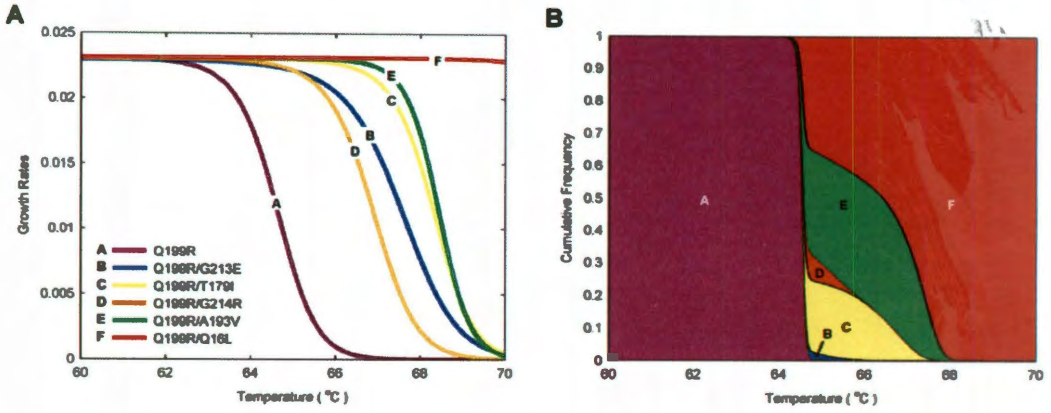


Figure 7.3
Fitness function modeled as a Hill function can reproduce experimental evolutionary dynamics. (A) Growth rates are calculated using the Hill-function (Figure 7.2) and *in vitro* enzyme activity between 60 and 70 °C (Figure 7.1B). (B) Growth rates in Figure 7.3A (along with mutation rate and seeding subpopulation sizes) are used in a simulation modeling growth of a clonal population to determine potential evolutionary outcomes, which correspond to experimental observations.

7.3 A Hill-like fitness function is consistent with Metabolic Flux Analysis

What is the influence of a single enzyme on the fitness of an organism? In trying to answer this question, Kacser and Burns along with Heinrich and Rapoport (Kacser & Burns, 1973; Heinrich & Rapoport, 1974b; Kacser & Burns, 1981) established the field of Metabolic Control Analysis whereby the fluxes through metabolic networks are quantitatively determined by identifying the structure and variables defining the network. Bridging flux to fitness provided the connections that allow the molecular attributes of a protein, like AK, to be linked to the properties of the metabolic systems under selection (Marx *et al.*, 2005). Resource-limited chemostat studies confirmed that growth rates

(fitness) were dependent on flux through the metabolic system as determined by the respective enzyme activities (Dean *et al.*, 1986; Dykhuizen *et al.*, 1987). The sensitivity of the flux (J) through a system to small changes in enzyme concentration (e_i) is described by the Flux Control Coefficient (C_i^J).

$$\frac{dJ}{J} \Big/ \frac{de_i}{e_i} = C_i^J \qquad 7\text{-}2$$

A consequence of C_i^J is a pattern of diminishing returns in flux with increasing enzyme concentration, a finding that has been confirmed in many studies (Salter *et al.*, 1986; Torres *et al.*, 1986; Dykhuizen *et al.*, 1987). The dependence of flux on activity or concentration of a single enzyme within a metabolic pathway are identical (e.g. an enzyme variant may have half the activity as wild type, or the system has half the enzyme concentration, the resulting effect on flux is the same regardless). By examining the extremes of the flux through metabolic networks, it is possible to appreciate the relationship of AK activity to cellular fitness. Without AK, the cell cannot maintain a sufficient energy charge and dies. At the other extreme, increasing activity cannot translate to faster growth rates because a series of metabolic pathways become less sensitive to modulations as enzyme concentration or activity increases ($C_i^J \rightarrow 0$). Simply put, increases to AK activity within the complex metabolic network reaches a point of diminishing returns. Understanding the dependence of fitness on activity between these extremes is critical to predicting evolutionary outcomes.

The simplified rate laws described in Figure 7.4 demonstrate that the response of ATP flux through the metabolic pathway saturates at high AK activity, resulting in the 'physiological threshold' suggested by our experimental evolution data

(Figure 7.5B). The differential equations used to describe the response of equilibrium concentrations of adenylate moieties, *M*, *D*, and *T* for AMP, ADP, and ATP respectively, to changes in activity (*a*) of AK are:

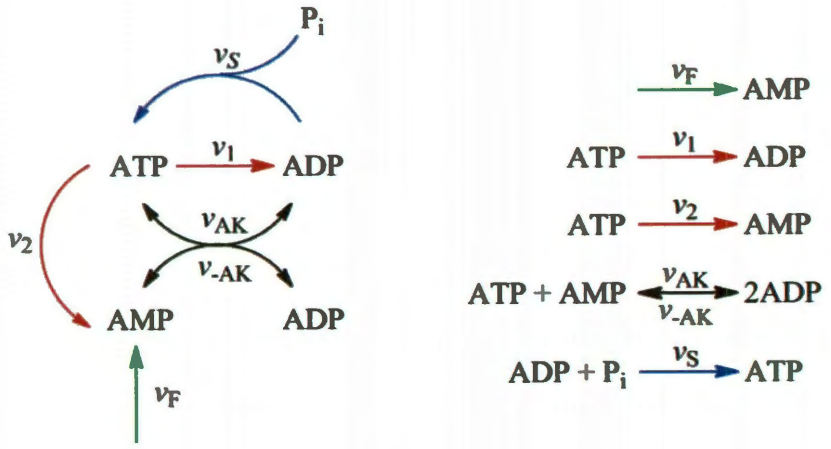


Figure 7.4
An idealized pathway for adenylate generation and use within a cell. AK reaction (v_{AK}, v_{-AK}), ATP synthase reaction (v_S), AMP formation (v_F), and ATP hydrolysis pathways to lower energy moieties (v_1, v_2) describe the dynamics of energy metabolism.

$$0 = \begin{cases} \dot{M} = \beta v_F - a v_{AK} MT + a v_{-AK} D^2 + v_2 T - \gamma M \\ \dot{D} = v_1 T + a v_{AK} MT - a v_{-AK} D^2 - v_S P_i D - \gamma D \\ \dot{T} = -a v_{AK} MT + a v_{-AK} D^2 - v_1 T - v_2 T + v_S P_i D - \gamma T \end{cases} \quad 7\text{-}3$$

Feedback exists in this system in a variety of forms and has been observed in a number of studies. *In vivo*, a high energy charge and ATP pool size correspond to growth rates (Glembotski *et al.*, 1981). ATP serves as a fundamental building block or energy source in nearly every aspect of cellular growth and replication. Faster growth increases utilization of ATP and increases demand for ATP through the ATP synthase pathway. The growth rate dependent demands on AK can be modeled as sinks of adenylate

moieties from the system (γ) as well as potentially increasing the flux of AMP into the system (β) with varying dependencies (n_β and n_γ).

$$\beta = n_\beta \frac{EC}{EC + EC_{max}} = n_\beta \frac{EC}{EC+1} \text{ and } \gamma = n_\gamma \frac{EC}{EC+1} \quad 7\text{-}4$$

Temperature will also affect rates, but these cases are ignored with the exception of the temperature dependence of AK activity. Our model suggests that energy metabolism catastrophically fails as enzyme activity becomes sufficiently low (Figure 7.5A). This corresponds to an ATP concentration and EC that approaches zero. The finding that a diminishing EC results in a catastrophic failure of energy metabolism is supported by *in vivo* observations for *E. coli* (Glembotski *et al.*, 1981) as well as in *G. stearothermophilus* (Chapman *et al.*, 1971). In *E. coli*, a high EC is maintained in most conditions, except for those in which cellular viability in severely limited (Glembotski *et al.*, 1981). Thus, EC is a critical indicator of viability. In growing recombinant *G. stearothermophilus*, the EC was experimentally determined to be ~0.23 at 55 °C (Couñago & Shamoo, 2005). A transient increase in temperature from 50 to 65 °C resulted in an EC of ~0.09 where an EC could not be determined after 30 seconds of incubation at 65 °C due to a high concentration of adenylate moieties in the media. This contrasts to an increase in EC (~0.4) after the same incubation period for the wild type, non-recombinant, strain. Modeling this system with seeding values reveals that the response of the EC in the system to a change in AK activity is indeed well approximated by a Hill function (Figure 7.5B). A more realistic model relating activity to flux and flux to fitness will ultimately lead to more accurate predictions of evolutionary outcomes and

demonstrates the power of modeling nonlinear systems to understand the complex and ubiquitous phenomenon of natural selection.

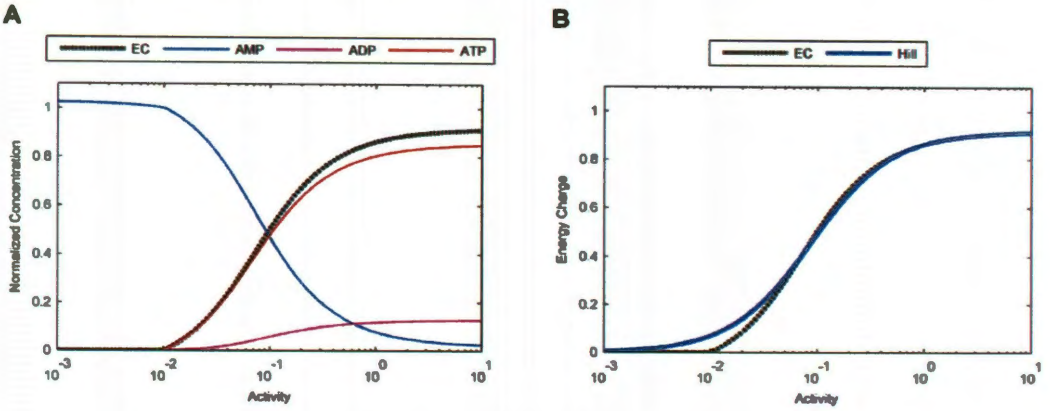


Figure 7.5
The response of energy charge to variation in AK activity can be modeled by a Hill function. (A) Concentrations of adenylate moieties and energy charge as a function of enzyme activity. Maximal energy charge would correspond to a system whereby growth rate is not limited by the ATP regeneration pathway. (B) The Hill function (blue) closely approximates the fall of EC as predicted by Metabolic Control theory (black dashed) that would occur as AK failed to maintain adenylate homeostasis.

7.4 Discussion

Experimental evolution of microbial populations has been used successfully to study adaptation to various pressures, from the presence of antibiotics to the absence of nutrients (Elena & Lenski, 2003; Dean & Thornton, 2007). In order for quantitative prediction of evolutionary outcomes to have success, we must have experimental systems that tightly couple the effect of mutations within a cellular population to their evolutionary fates. The 'weak link' approach creates a node within a metabolic network that has a large flux control coefficient ($C_i^J \rightarrow 1$), meaning that improvements to the weak link will have the greatest impact on fitness. This focuses adaptation during selection

primarily to this single gene. This has the benefit of simplifying the identification of mutations and allows for the study of changes in properties of a single enzyme to fitness. Most importantly, adaptation through a single gene can capture many of the complexities and nuances of the larger system, including nonlinear phenomena such as mutational epistasis and saturation effects due to large-scale metabolic networks.

While much has been made of identifying and characterizing mutations as candidates of adaptation, less is known of how these mutations correspond to alterations of protein structure and function that ultimately alter fitness and shape molecular evolution. Metabolic Control Analysis has been used to determine fitness of organisms over a range of enzyme expression levels but the theory has yet to be combined with models of bacterial growth and evolution to predict a population's response to stress. By correlating fitness to enzyme activity and population dynamics to point mutations, we have begun to understand the multi-scale phenomena driving evolution. This same approach can be applied to the study of other evolutionary processes – like the emergence of antibiotic resistance in pathogens or the adaptation of cancer cells to chemotherapy – thereby aiding in the design and implementation of treatment strategies.

The successful implementation of models relating physicochemical properties to evolutionary dynamics permits hypothesis driven ideas to undergo robust evaluation. Biophysical studies of enzyme variants isolated from adapted intermediates during experimental evolution effectively reveal the enzyme properties that are conserved or enhanced by natural selection. Although the adaptive mutations to AK alter stability and activity through complex mechanisms at the molecular level, each isolated variant extends essential activity to higher temperatures which translates to increased range of

viability (Miller *et al.*, 2010; Peña *et al.*, 2010a). Flux based models provide general and scalable systems for linking *in vitro* properties of molecules like AK to fitness and evolutionary dynamics thereby bridging intra- and intermolecular influences. Most importantly, these systems afford an opportunity to critically evaluate which of many measurable properties are most critical for predictive success. In the case of the thermal adaptation of AK, classic biochemical properties such as catalytic rate (k_{cat}) and the Michaelis constant (K_M) must be combined with a thorough evaluation of AK stability and folding for quantitative success to be achieved. The 'weak link' approach provides a critically needed system for honing our models for adaptation and moving towards a more quantitative understanding of natural selection with important implications for biotechnology and biomedicine.

Chapter 8 – Results and Implications

8.1 Summary of discoveries and unique contributions

The model developed herein that relates adenylate kinase activity to fitness in a temperature sensitive strain represents the first use of *in vitro* enzyme activity within the framework of a fitness function to accurately simulate evolutionary outcomes. We are limited to predicting outcomes of a competition based on the properties of variants that have been carefully characterized. The folding components, both reversible and irreversible, are required to accurately interpret the catalytic properties of AK variants and to determine the fraction of the ensemble that is properly folded. Accounting for folding processes was necessary for understanding adaption in a model system where enzyme function is being challenged at increasingly higher temperatures. The folding component may be negligible in other systems where variants aren't substantially unfolded. However, many adaptive pathways, especially those that alter specificity towards substrates, as is the case for enzymes that confer extended-spectrum antibiotic resistance, contain global suppressor mutations that counteract mutations that alter function at the cost of stability (Brown *et al.*). Thus, this model could be applied more broadly to adaptive pathways distinguished by changing specificity and corresponding decreases in enzyme stability.

The model captures the relationship between AK activity at various temperatures and the corresponding relative fitness among strains carrying variants. However, no model is complete, and some work could be done to improve various aspects, some of which I will describe here briefly. A maximum, invariable growth rate was applied to

strains within the simulation with enzymes above the threshold (see Figure 7.3A), but the growth rate varies for wild type as a function of temperature when AK activity is not limiting (Figure 6.5). It is well known that microorganisms have optimal growth temperatures, and yet the current model cannot reproduce this phenomenon. A simple solution would be to use the growth-temperature profile of the host organism as a constraint on the maximal growth rate. This assumes that AK activity in a recombinant strain cannot improve fitness above that of the host with its endogenous *G. stearothermophilus adk* allele.

Secondly, the model suggests that aggregation is related to fitness to the extent that it depletes the pool of recycled enzymes available to perform an essential function. This model does not account for the intrinsic cytotoxic effect of misfolding, and we have made it clear that misfolding or aggregation in our system is analogous to but not a measure of *in vivo* misfolding. The exponential loss of enzyme suggests irreversible denaturation occurring through a Lumry-Eyring model is more prominent under conditions where the protein is more unstable or has a greater tendency to misfold (see Figure 7.1 for experimental evidence of this). Therefore, to reduce the *in vivo* cost of misfolding, natural selection could "remedy" the fitness cost by selecting for a more stable protein (leading to a lower concentration of the unfolded state which serves as an intermediate on the pathway to the denatured state) or reduce the tendency to misfold from the unfolded state (decrease k_{irrev}). Recent work has quantified the fitness cost of misfolding (Geiler-Samerotte *et al.*, 2011). However, it is not clear whether one mechanism, either improved stability through suppressor mutations or a decreased tendency to misfold, or both serve as predominant corrections to this.

One strength of the model is that it has been used to understand what were once puzzling results. The 'reverse experiment', using AK_{BSUB}Q199R/Q16L strain as a parental strain and decreasing temperature as a function of time, did not yield any new mutations. At an earlier time, it was expected that a destabilizing mutant would occur making the enzyme less rigid and thus more catalytically efficient at lower temperatures. However, no mutations were seen, suggesting that there was no benefit for such mutations. This was not expected due to previous measurements which showed Q199R as having a greater fitness relative to Q199R/Q16L at 55 °C (Couñago *et al.*, 2006). In light of the model with an activity threshold this outcome is not surprising and would instead be expected. The growth rate assay performed on strains containing AK_{BSUB}Q199R, AK_{BSUB}Q199R/Q16L and AK_{GSTE}, which do not agree with previous results, demonstrates that there is no fitness cost carried by the Q199R/Q16L allele as temperature decreases.

8.2 Future directions

The system and the activity-fitness model that has been developed could be used to answer specific questions regarding differences in stability measured between *in vitro* and *in vivo* environments as well as being used to probe how the demands of the cell change as a function of growth rate and temperature. The fact that the *B. subtilis adk* Q199R allele confers viability beyond 60 °C and yet has a melting temperature of approximately 50 °C is strong evidence that stabilizing effects exist in the intracellular milieu. Here I suggest that substrate can confer a significant level of stabilization, but many other mechanisms are known to buffer against instability (chaperones and crowding). What is not known is the relative influence of each *in vivo*. Knocking out

chaperones in the NUB3621-R:ThEv system may reveal their specific impact in distinct decrease in fitness for each of the mutant strains.

The relation between fitness and activity in this system is treated as a constant, but this need not be the case. One can imagine that as the organism reaches its optimal growth temperature, turnover of ATP increases as more is being used in the synthesis of peptides, RNA, etc. The threshold, the activity required to satisfy metabolic demands, increases as you approach the optimal temperature (as is suggested on page 112). Competition between pairs of mutants can be used to approximate the threshold as a function of temperature. One would need to determine the temperature at which two mutants have equivalent growth rates. This would identify a point where one enzyme variant is at the threshold, and the other above it. At higher temperatures, the growth rates would diverge signifying that the activity of one has fallen below the threshold. This experiment identifies the threshold at a specific temperature. Comparison between multiple pairs of mutants could be used to determine the threshold and provide validation for how this threshold changes with temperature.

We can reconcile evolutionary outcomes with *in vitro* (or *in vivo*) kinetic data. However, in the future we should be able to predict the site of adaptations before they happen. Modern techniques for studying fitness effects in microbial strains where each gene is knocked out or its expression amplified can provide direction with where to look (Nichols *et al.*, 2011). If fitness in the presence of a drug is severely affected by the absence or excess of gene product compared to wild type then it suggests the gene is a constituent in a pathway for drug degradation, removal, or displacement. This information can be used to piece together new pathways or complete known partial

pathways. Metabolic control analysis can then be used to identify the existing weak links once the pathway is constructed (enzymes with a high flux control coefficient) and monitored for adaptation. As of now, many targets of drug resistance cannot be predicted *a priori*. However, tools are being developed to better describe the networks, which in turn is essential to prediction of critical metabolic nodes.

It would also be worth investigating how metabolic systems respond to a continually increased stress, as was done for the weak link, but for systems where metabolic control is more evenly distributed across multiple enzymatic links. The characteristic strong dependence of fitness on AK activity is what made the system well behaved with respect to the experimenter's expectations. However, most systems do not have a single control element. What will be the dynamics of changes, regulatory or adaptive, observed during the optimization of a system to an external selective pressure that perturbs either kinetics of metabolic links or their pools of substrates?

The complexity of biological systems leaves much to be understood. Adaptation provides us with a way of perturbing these systems so that we many understand how mutations influence enzymes and how enzymes influence the organism. Theodosius Dobzhansky had written a series of essays with the often quoted title *Nothing in biology makes sense except in light of evolution*. Based on the work in this thesis, the works on which it was built, and the work that will surely come, it could be said that "Nothing in evolution makes sense except in light of biochemistry."[29]

[29] Quoted from Dr. Yousif Shamoo, a great advisor and friend.

Bibliography

Atkinson, D.E. (1966). Regulation of enzyme activity. Annu Rev Microbiol *35*, 85-124.

Atkinson, D.E. (1968). The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers. Biochemistry *7*, 4030-4034.

Atkinson, D.E. (1969). Regulation of enzyme function. Annu Rev Microbiol *23*, 47-68.

Atkinson, D.E. (1977). Cellular energy metabolism and its regulation (New York, Academic Press, Inc.).

Atkinson, D.E., and Walton, G.M. (1967). Adenosine triphosphate conservation in metabolic regulation. Rat liver citrate cleavage enzyme. J Biol Chem *242*, 3239-3241.

Bae, E., and Phillips, G.N., Jr. (2004). Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. J Biol Chem *279*, 28202-28208.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with escherichia coli. Nature *461*, 1243-1247.

Barzu, O., and Michelson, S. (1983). Simple and fast purification of escherichia coli adenylate kinase. FEBS Lett *153*, 280-284.

Bateson, W., and Mendel, G. (1909). Mendel's principles of heredity (Cambridge [Eng.], At the University Press).

Becktel, W.J., and Schellman, J.A. (1987). Protein stability curves. Biopolymers *26*, 1859-1877.

Bennett, A.F., Dao, K.M., and Lenski, R.E. (1990). Rapid evolution in response to high-temperature selection. Nature *346*, 79-81.

Bèrg, J., Hung, Y.P., and Yellen, G. (2009). A genetically encoded fluorescent reporter of atp:Adp ratio. Nat Methods *6*, 161-166.

Bershtein, S., Goldin, K., and Tawfik, D.S. (2008). Intense neutral drifts yield robust and evolvable consensus proteins. J Mol Biol *379*, 1029-1044.

Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C., and Arnold, F.H. (2005). Thermodynamic prediction of protein neutrality. Proc Natl Acad Sci U S A *102*, 606-611.

Bochner, B.R., and Ames, B.N. (1982). Complete analysis of cellular nucleotides by two-dimensional thin layer chromatography. J Biol Chem *257*, 9759-9769.

Boveri, T. (1902). Über mehrpolige mitosen als mittel zur analyse des zellkerns. Verh d Phys Med Ges zu Würzburg
Verhandlungen der physicalisch-medizinischen Gesselschaft zu Würzburg *35*, 67-90.

Brandts, J.F., and Lin, L.N. (1990). Study of strong to ultratight protein interactions using differential scanning calorimetry. Biochemistry *29*, 6927-6940.

Briggs, G.E., and Haldane, J.B. (1925). A note on the kinetics of enzyme action. Biochem J *19*, 338-339.

Brown, K.M., Costanzo, M.S., Xu, W., Roy, S., Lozovsky, E.R., and Hartl, D.L. (2010). Compensatory mutations restore fitness during the evolution of dihydrofolate reductase. Mol Biol Evol *27*, 2682-2690.

Brown, N.G., Pennington, J.M., Huang, W., Ayvaz, T., and Palzkill, T. Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum tem beta-lactamases. J Mol Biol *404*, 832-846.

Buckstein, M.H., He, J., and Rubin, H. (2008). Characterization of nucleotide pools as a function of physiological state in escherichia coli. J Bacteriol *190*, 718-726.

Bull, J., and Wichman, H. (1998). A revolution in evolution. Science *281*, 1959.

Bull, J.J., Badgett, M. R., Rokyta, D., and Molineux, I. J. (2003). Experimental evolution yields hundreds of mutations in a functional viral genome. J Mol Evol *57*.

Burch, C.L., and Chao, L. (1999). Evolution by small steps and rugged landscapes in the rna virus phi6. Genetics *151*, 921-927.

Chapman, A.G., and Atkinson, D.E. (1977). Adenine nucleotide concentrations and turnover rates. Their correlation with biological activity in bacteria and yeast. In Advances in microbial physiology, A.H. Rose, and D.W. Tempest, eds. (London, Academic Press), pp. 253-306.

Chapman, A.G., Fall, L., and Atkinson, D.E. (1971). Adenylate energy charge in escherichia coli during growth and starvation. J Bacteriol *108*, 1072-1086.

Chen, P., and Shakhnovich, E.I. (2009). Lethal mutagenesis in viruses and bacteria. Genetics *183*, 639-650.

Chen, Z.F., Wojcik, S.F., and Welker, N.E. (1986). Genetic analysis of bacillus stearothermophilus by protoplast fusion. J Bacteriol *165*, 994-1001.

Cleland, W.W. (1963). The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations. Biochim Biophys Acta *67*, 104-137.

Cooper, T.F., Rozen, D.E., and Lenski, R.E. (2003). Parallel changes in gene expression after 20,000 generations of evolution in escherichiacoli. Proc Natl Acad Sci U S A *100*, 1072-1077.

Copley, S.D. (2009). Evolution of efficient pathways for degradation of anthropogenic chemicals. Nat Chem Biol *5*, 559-566.

Cornish-Bowden, A. (1979). Fundamentals of enzyme kinetics 1edn (London and Boston, Buttersworths).

Correns, C. (1900). G. Mendel's regel über das verhalten der nachkommenschaft der rassenbastarde. Berichte der deutschen botanischen Gesellschaft *18*, 158-168.

Couñago, R., Chen, S., and Shamoo, Y. (2006). In vivo molecular evolution reveals biophysical origins of organismal fitness. Mol Cell *22*, 441-449.

Couñago, R., and Shamoo, Y. (2005). Gene replacement of adenylate kinase in the gram-positive thermophile geobacillus stearothermophilus disrupts adenine nucleotide homeostasis and reduces cell viability. Extremophiles *9*, 135-144.

Couñago, R., Wilson, C.J., Peña, M.I., Wittung-Stafshede, P., and Shamoo, Y. (2008). An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability-activity trade-offs. Protein Eng Des Sel *21*, 19-27.

Cousin, D., and Belaich, J.P. (1966). On a thermosensitive mutation of escherichia coli affecting an energizing function. C R Acad Sci Hebd Seances Acad Sci D *263*, 886-888.

Creighton, T.E. (1983). Proteins : Structures and molecular principles (New York, W.H. Freeman).

Creighton, T.E. (1988). Toward a better understanding of protein folding pathways. Proc Natl Acad Sci U S A *85*, 5082-5086.

Darwin, C. (1859). On the origin of species (John Murray).

Darwin, C., and Wallace, A.R. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. Journal of the Proceedings of the Linnean Society of London *3*, 46-50.

De Beer, G. (1964). Mendel, darwin, and fisher (1865-1965). Notes Rec R Soc Lond *19*, 192-226.

De Vries, H. (1900). Sur la loi de disjonction des hybrides. Comptes Rendus de l'Academie des Sciences *130*, 845-847.

Dean, A.M. (1995). A molecular investigation of genotype by environment interactions. Genetics *139*, 19-33.

Dean, A.M., Dykhuizen, D.E., and Hartl, D.L. (1986). Fitness as a function of beta-galactosidase activity in escherichia coli. Genet Res *48*, 1-8.

Dean, A.M., and Thornton, J.W. (2007). Mechanistic approaches to the study of evolution: The functional synthesis. Nat Rev Genet *8*, 675-688.

Dill, K.A. (1990). Dominant forces in protein folding. Biochemistry *29*, 7133-7155.

Drake, J.W., Charlesworth, B., Charlesworth, D., and Crow, J.F. (1998). Rates of spontaneous mutation. Genetics *148*, 1667-1686.

Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell *134*, 341-352.

Dykhuizen, D.E., Dean, A.M., and Hartl, D.L. (1987). Metabolic flux and fitness. Genetics *115*, 25-31.

Elena, S.F., and Lenski, R.E. (2003). Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. Nat Rev Genet *4*, 457-469.

Esmon, B.E., Kensil, C.R., Cheng, C.H., and Glaser, M. (1980). Genetic analysis of escherichia coli mutants defective in adenylate kinase and sn-glycerol 3-phosphate acyltransferase. J Bacteriol *141*, 405-408.

Fisher, R.A. (1928). The possible modification of the response of the wild type to recurrent mutations. The American Naturalist *62*, 115-126.

Fisher, R.A. (1930). The genetical theory of natural selection (Oxford,, The Clarendon Press).

Freire, E., and Biltonen, R.L. (1978). Statistical mechanical deconvolution of thermal transitions in macromolecules. I. Theory and
application to homogeneous systems. Biopolymers *17*, 463-479.

Geiler-Samerotte, K.A., Dion, M.F., Budnik, B.A., Wang, S.M., Hartl, D.L., and Drummond, D.A. (2011). Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proc Natl Acad Sci U S A *108*, 680-685.

Gilles, A.M., Saint-Girons, I., Monnot, M., Fermandjian, S., Michelson, S., and Barzu, O. (1986). Substitution of a serine residue for proline-87 reduces catalytic activity and increases susceptibility to proteolysis of escherichia coli adenylate kinase. Proc Natl Acad Sci U S A *83*, 5798-5802.

Glaser, M., Bayer, W.H., Bell, R.M., and Vagelos, P.R. (1973). Regulation of macromolecular biosynthesis in a mutant of escherichia coli defective in membrane phospholipid biosynthesis. Proc Natl Acad Sci U S A *70*, 385-389.

Glaser, M., Nulty, W., and Vagelos, P.R. (1975). Role of adenylate kinase in the regulation of macromolecular biosynthesis in a putative mutant of escherichia coli defective in membrane phospholipid biosynthesis. J Bacteriol *123*, 128-136.

Glaser, P., Presecan, E., Delepierre, M., Surewicz, W.K., Mantsch, H.H., Barzu, O., and Gilles, A.M. (1992). Zinc, a novel structural element found in the family of bacterial adenylate kinases. Biochemistry *31*, 3038-3043.

Glembotski, C.C., Chapman, A.G., and Atkinson, D.E. (1981). Adenylate energy charge in escherichia coli cr341t28 and properties of heat-sensitive adenylate kinase. J Bacteriol *145*, 1374-1385.

Gould, S.J. (1989). Wonderful life: The burgess shale and the nature of history (New York, Norton).

Haase, G.H., Brune, M., Reinstein, J., Pai, E.F., Pingoud, A., and Wittinghofer, A. (1989). Adenylate kinases from thermosensitive escherichia coli strains. J Mol Biol *207*, 151-162.

Haldane, J.B.S. (1930). A note on fisher's theory of the origin of dominance, and on a correlation between dominance and linkage. The American Naturalist *64*, 87-90.

Hall, B.G. (2002). Predicting evolution by in vitro evolution requires determining evolutionary pathways. Antimicrob Agents Chemother *46*, 3035-3038.

Hartl, D.L., Dykhuizen, D.E., and Dean, A.M. (1985). Limits of adaptation: The evolution of selective neutrality. Genetics *111*, 655-674.

Heinrich, R., and Rapoport, T.A. (1974a). A linear steady-state treatment of enzymatic chains. Critique of the crossover theorem and a general procedure to identify interaction sites with an effector. Eur J Biochem *42*, 97-105.

Heinrich, R., and Rapoport, T.A. (1974b). A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur J Biochem *42*, 89-95.

Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. Nature *450*, 964-972.

Henzler-Wildman, K.A., Lei, M., Thai, V., Kerns, S.J., Karplus, M., and Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature *450*, 913-916.

Hoff, K.G., Culler, S.J., Nguyen, P.Q., McGuire, R.M., Silberg, J.J., and Smolke, C.D. (2009). In vivo fluorescent detection of fe-s clusters coordinated by human grx2. Chem Biol *16*, 1299-1308.

Huxley, J. (1942). Evolution, the modern synthesis (London,, G. Allen & Unwin ltd).

Jackson, S.E. (1998). How do small single-domain proteins fold? Fold Des *3*, R81-91.

Kacser, H., and Beeby, R. (1984). Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection. J Mol Evol *20*, 38-51.

Kacser, H., and Burns, J.A. (1973). The control of flux. Symp Soc Exp Biol *27*, 65-104.

Kacser, H., and Burns, J.A. (1979). Molecular democracy: Who shares the controls? Biochem Soc Trans *7*, 1149-1160.

Kacser, H., and Burns, J.A. (1981). The molecular basis of dominance. Genetics *97*, 639-666.

Knies, J.L., Izem, R., Supler, K.L., Kingsolver, J.G., and Burch, C.L. (2006). The genetic basis of thermal reaction norm evolution in lab and natural phage populations. PLoS Biol *4*, e201.

Krebs, H. (1964). The croonian lecture, 1963. Gluconeogenesis. Proc R Soc Lond B Biol Sci *159*, 545-564.

Lehninger, A., Nelson, D.L., and Cox, M.M. (2008). Lehninger principles of biochemistry, Fifth edn (New York, W. H. Freeman and Company).

Lenski, R.E., and Travisano, M. (1994). Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. Proc Natl Acad Sci U S A *91*, 6808-6814.

Leskovac, V. (2003). Comprehensive enzyme kinetics (New York, Kluwer Academic/Plenum Pub.).

Lumry, R., and Eyring, H. (1954). Conformation changes of proteins. J Phys Chem *58*, 110-120.

Lunzer, M., Miller, S.P., Felsheim, R., and Dean, A.M. (2005). The biochemical architecture of an ancient adaptive landscape. Science *310*, 499-501.

Marx, C.J., Van Dien, S.J., and Lidstrom, M.E. (2005). Flux analysis uncovers key role of functional redundancy in formaldehyde metabolism. PLoS Biol *3*, e16.

Mendel, G. (1866). Versuche über pflanzen-hybriden. Verh Naturforsch Ver Brünn *4*, 3-47.

Michaelis, L., and Menten, M.L. (1913). Kinetik der invertinwirkung. Biochemisches Zeitschrift *49*, 333-369.

Miller, C., Davlieva, M., Wilson, C., White, K.I., Couñago, R., Wu, G., Myers, J.C., Wittung-Stafshede, P., and Shamoo, Y. (2010). Experimental evolution of adenylate kinase reveals contrasting strategies toward protein thermostability. Biophys J *99*, 887-896.

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. Nature *191*, 144-148.

Nei, M., and Koehn, R.K. (1983). Evolution of genes and proteins (Sunderland, Mass., Sinauer Associates).

Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A., *et al.* (2011). Phenotypic landscape of a bacterial cell. Cell *144*, 143-156.

Orr, H.A. (2003). The distribution of fitness effects among beneficial mutations. Genetics *163*, 1519-1526.

Orr, H.A. (2009). Fitness and its role in evolutionary genetics. Nat Rev Genet *10*, 531-539.

Pace, C.N. (1975). The stability of globular proteins. CRC Crit Rev Biochem *3*, 1-43.

Pace, C.N., and McGrath, T. (1980). Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation. J Biol Chem *255*, 3862-3865.

Peña, M.I., Davlieva, M., Bennett, M.R., Olson, J.S., and Shamoo, Y. (2010a). Evolutionary fates within a microbial population highlight an essential role for protein folding during natural selection. Mol Syst Biol *6*, 387.

Peña, M.I., Van Itallie, E., Bennett, M.R., and Shamoo, Y. (2010b). Evolution of a single gene highlights the complexity underlying molecular descriptions of fitness. Chaos *20*, 026107.

Pisliakov, A.V., Cao, J., Kamerlin, S.C., and Warshel, A. (2009). Enzyme millisecond conformational dynamics do not catalyze the chemical step. Proc Natl Acad Sci U S A *106*, 17359-17364.

Privalov, P.L. (1979). Stability of proteins: Small globular proteins. Adv Protein Chem *33*, 167-241.

Radika, K., and Northrop, D.B. (1984). Correlation of antibiotic resistance with vmax/km ratio of enzymatic modification of aminoglycosides by kanamycin acetyltransferase. Antimicrob Agents Chemother *25*, 479-482.

Rhoades, E., Gussakovsky, E., and Haran, G. (2003). Watching proteins fold one molecule at a time. Proc Natl Acad Sci U S A *100*, 3197-3202.

Rhoads, D.G., and Lowenstein, J.M. (1968). Initial velocity and equilibrium kinetics of myokinase. J Biol Chem *243*, 3963-3972.

Robertson, A.D., and Murphy, K.P. (1997). Protein structure and the energetics of protein stability. Chem Rev *97*, 1251-1268.

Rutherford, S.L. (2003). Between genotype and phenotype: Protein chaperones and evolvability. Nat Rev Genet *4*, 263-274.

Saint Girons, I., Gilles, A.M., Margarita, D., Michelson, S., Monnot, M., Fermandjian, S., Danchin, A., and Barzu, O. (1987). Structural and catalytic characteristics of escherichia coli adenylate kinase. J Biol Chem *262*, 622-629.

Salter, M., Knowles, R.G., and Pogson, C.I. (1986). Quantification of the importance of individual steps in the control of aromatic amino acid metabolism. Biochem J *234*, 635-647.

Sanchez-Ruiz, J.M. (1992). Theoretical analysis of lumry-eyring models in differential scanning calorimetry. Biophys J *61*, 921-935.

Sanchez-Ruiz, J.M., Lopez-Lacomba, J.L., Cortijo, M., and Mateo, P.L. (1988). Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. Biochemistry *27*, 1648-1652.

Sasaki, M., and Kurusu, Y. (2004). Analysis of spontaneous base substitutions generated in mutator strains of bacillus subtilis. FEMS Microbiol Lett *234*, 37-42.

Schellman, J.A. (1975). Macromolecular binding. Biopolymers *14*, 999-1018.

Schneider, D.A., and Gourse, R.L. (2004). Relationship between growth rate and atp concentration in escherichia coli: A bioassay for available cellular atp. J Biol Chem *279*, 8262-8268.

Schrank, T.P., Bolen, D.W., and Hilser, V.J. (2009). Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins. Proc Natl Acad Sci U S A *106*, 16984-16989.

Soskine, M., and Tawfik, D.S. (2010). Mutational effects and the evolution of new protein functions. Nat Rev Genet *11*, 572-582.

Sutton, W.S. (1902). On the morphology of the chromosome group in brachystola magna. Biological Bulletin *4*, 24-39.

Thomas, T.M., and Scopes, R.K. (1998). The effects of temperature on the kinetics and stability of mesophilic and thermophilic 3-phosphoglycerate kinases. Biochem J *330 (Pt 3)*, 1087-1095.

Tokuriki, N., and Tawfik, D.S. (2009a). Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature *459*, 668-673.

Tokuriki, N., and Tawfik, D.S. (2009b). Stability effects of mutations and protein evolvability. Curr Opin Struct Biol.

Torres, N.V., Mateo, F., Melendez-Hevia, E., and Kacser, H. (1986). Kinetics of metabolic pathways. A system in vitro to study the control of flux. Biochem J *234*, 169-174.

Travisano, M., Mongold, J.A., Bennett, A.F., and Lenski, R.E. (1995). Experimental tests of the roles of adaptation, chance, and history in evolution. Science *267*, 87-90.

Tschermak, E. (1900). Über künstliche kreuzung bei pisum sativum. Zeilschrifl Jur das landwirthschaftliche Versuchswesen in Oeslerreich *3*, 465-555.

Turner, P.E., Burch, C.L., Hanley, K.A., and Chao, L. (1999). Hybrid frequencies confirm limit to coinfection in the rna bacteriophage phi6. J Virol *73*, 2420-2424.

Wakeley, J. (1994). Substitution-rate variation among sites and the estimation of transition bias. Mol Biol Evol *11*, 436-442.

Wakeley, J. (1996). The excess of transitions among nucleotide substitutions: New methods of estimating transition bias underscore its significance. Trends in Ecology & Evolution *11*, 158-162.

Wichman, H.A., Badgett, M.R., Scott, L.A., Boulianne, C.M., and Bull, J.J. (1999). Different trajectories of parallel evolution during viral adaptation. Science *285*, 422-424.

Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E.Z., and Kern, D. (2004). Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat Struct Mol Biol *11*, 945-949.

Wright, S. (1929). Fisher's theory of dominance. The American Naturalist *63*, 274-279.

Wright, S. (1934). Physiological and evolutionary theories of dominance. The American Naturalist *68*, 24-53.

Wu, L.J., and Welker, N.E. (1989). Protoplast transformation of bacillus stearothermophilus nub36 by plasmid DNA. J Gen Microbiol *135*, 1315-1324.

Zeldovich, K.B., Chen, P., and Shakhnovich, E.I. (2007). Protein stability imposes limits on organism complexity and speed of molecular evolution. Proc Natl Acad Sci U S A *104*, 16152-16157.

Zwanzig, R. (1997). Two-state models of protein folding kinetics. Proc Natl Acad Sci U S A *94*, 148-150.

Appendix

A.1 Specialized Media

Medium	Component	Amount or Concentration
Modified LB (mLB)	Tryptone	10 g
1 L	Yeast Extract	5 g
	NaCl	5 g
	Agar (if desired)	15 g
	Water	995 mL
	NaOH (10% w/v)	1.25 mL
	Autoclave, and cool to 55 °C and add 1 mL from each of the following sterile stocks.	
	Nitrilotriacetic acid	1.05 M
	$MgSO_4 \cdot 7H_2O$	0.59 M
	$CaCl_2 \cdot 2H_2O$	0.91 M
	$FeSO_4 \cdot H_2O$	0.04 M
TS	Tryptone	40 g
1 L	NaCl	5 g
	Agar (if desired)	15
	Water	995 mL
	NaOH (10% w/v)	1.25 mL
	Autoclave, and cool to 55 °C and add 1 mL from each of the following sterile stocks.	
	Nitrilotriacetic acid	1.05 M
	$MgSO_4 \cdot 7H_2O$	0.59 M
	$CaCl_2 \cdot 2H_2O$	0.91 M
	$FeSO_4 \cdot H_2O$	0.04 M
Modified Protoplasting (mP)	4X mLB	12.5 mL
50 mL	20% Lactose	25 mL
	1 M $MgCl_2 \cdot 6H_2O$	0.5 mL
	Sterile water	12 mL
Regenerative Agar (RA)[1]	Bacto Agar	4 g
0.5 L	Water	115 mL
	Autoclave, and cool to 55 °C and add aseptically add the following mixtures.	
	20% Lactose	250 mL
	4X mLB	125 mL
	1 M $MgCl_2 \cdot 6H_2O$	5 mL
	2 M $CaCl_2 \cdot 2H_2O$	5 mL

1 Prepare RA plates the day before or day of the transformation.

A.2 Plasmids

Plasmid	Relevant Genotype	Source / Reference
pSTE12	Amp^r, Tet^{r1}	Nakayama et al. (1992)
p12ThEv	Amp^r, Tet^r, adk_{Bsub}, Cml^r	Couñago and Shamoo (2005)
pSTE-promo	Amp^r, Tet^r, $P_{Gste\text{-}spc}$	Not described in this work
pSTE-Q199R/Q16L	Amp^r, Tet^r, $P_{Gste\text{-}spc}$, $adk_{B.sub\text{-}Q199R/Q16L}$	Not described in this work
pSTE-YFP	Amp^r, Tet^r, $P_{Gste\text{-}spc}$, yfp_{Venus}	Not described in this work
pSTE-(T7)YFP	Amp^r, Tet^r, $PG_{ste\text{-}spc}$, P_{T7}, yfp_{Venus}	Not described in this work
pSTE-mQ199R	Amp^r, Tet^r, $P_{Gst\text{-}spc}$, mutagenized adk_{Bsub}	Not described in this work
pET-11a	Amp^r, T7*lac*, *lacI*	Novagen (now EMD Chemicals, Inc., Gibbstown, NJ)
pET-Gstadk	Amp^r, Tet^r, T7*lac*, *lacI*, adk_{Gste}	This work
pTara	Cml^r, *araC*, P_{bad}, T7 RNAP	Wycuff and Matthews (2000)

1 The tetracycline resistance protein encoded in pSTE12 and derivatives is thermolabile, and selection at temperatures greater than 60 °C should be avoided.

A.3 Primers and applications

Primers used to amplify fragments used in the sub-cloning of novel constructs

Plasmid	Primers	Template	Polymerase	Thermocycler program
pSTE-YFP	Venus_SphI_F Venus_HindIII_R	Plasmid containing full length Venus (Silberg Lab)[1]	Phusion	Phusion
pSTE-(T7)YFP	T7_pSTE_F T7_pSTE_R	pSTE-YFP	PfuTurbo AD	Pfu insertion
pSTE-Q199R/Q16L	Bs_adk_SphI_F Bs_adk_HindIII_R	p12ThEv-Q199R	Phusion	G stearo adk
pSTE-promo	Promoter_SacI_F Promoter_SphI_R		Phusion	Phusion
pSTE-mQ199R (Madk)	Mut_Bs_adk_SphI_F Mut_Bs_adk_HindIII_R	PCR product	Mutazyme II	EP PCR
pET-Gstadk	NUB_ADK_F_NdeI NUB_ADK_R_BamHI	NUB3621-R genomic	Phire	G stearo adk

References:
1 (Hoff *et al.*, 2009)

A.4 Derivation of the modified Michaelis-Menten equation accounting for folding and irreversible misfolding

Fitness is assumed to be dependent on the level of intracellular adenylate kinase (AK) activity. AK has a random bi bi kinetic mechanism and catalyzes the reversible

reaction: ATP + AMP ⇄ 2ADP (Rhoads & Lowenstein, 1968). A quasi-steady state approximation for the instantaneous rates of ADP production *in vitro* by *B. subtilis* AK was used to fit the kinetic data.

$$\frac{d[\text{Product}]}{dt} = v = \frac{k_{cat}[S][E]_0}{K_M(1+K_{unf})+[S]} e^{-k_{irrev} \cdot t \cdot Y_U} \qquad \text{A-1}$$

$$Y_U = \frac{K_M K_{unf}}{K_M(1+K_{unf})+[S]} \qquad \text{A-2}$$

In Equation A-1, K_M is the Michaelis-Menten constant, k_{cat} is the catalytic rate, $[E]_0$ is the total enzyme, K_{unf} is the ratio of the unfolding rates (unfolding equilibrium), k_{irrev} is the rate of the irreversible loss of protein, $[S]$ is the substrate concentration, and Y_U is the fraction of unfolded protein. The rate equation is determined by inspection of the enzyme species and equilibrium constants. The Michaelis-Menten constant (K_M), unfolding equilibrium (K_{unf}), and total enzyme concentration ($[E]_0$) are:

$$K_M = \frac{[E][S]}{[ES]}$$

$$K_{unf} = \frac{[U]}{[E]}$$

$$[E]_0 = [E]+[ES]+[U]$$

where the U and E states are assumed to be in rapid equilibrium compared to the catalytic reactions. The expression for total enzyme concentration in terms of [ES], K_M, K_{unf}, and [S] is therefore:

$$[E]_0 = \frac{K_M[ES]}{[S]} + [ES] + \frac{K_M K_{unf}[ES]}{[S]}$$

$$[E]_0 = [ES]\left(\frac{K_M}{[S]} + 1 + \frac{K_M K_{unf}}{[S]}\right)$$

Rearranging the equation to provide a definition for [ES] and multiplying both sides by the catalytic rate constant (k_{cat}) yields an equation describing the initial reaction rate.

$$[ES] = \frac{[E]_0}{\left(\frac{K_M}{[S]} + 1 + \frac{K_M K_{unf}}{[S]}\right)}$$

$$[ES] = \frac{[E]_0[S]}{(K_M + [S] + K_M K_{unf})}$$

$$v_i = k_{cat}[ES] = \frac{k_{cat}[E]_0[S]}{(K_M + [S] + K_M K_{unf})} = \frac{k_{cat}[E]_0[S]}{K_M(1 + K_{unf}) + [S]}$$

The fraction of unfolded protein is defined as done previously for $[E]_0$.

$$Y_U = \frac{[U]}{[E]_0} = \frac{[U]}{[E] + [ES] + [U]}$$

$$Y_U = \frac{\frac{K_M K_{unf}[ES]}{[S]}}{\frac{K_M[ES]}{[S]} + [ES] + \frac{K_M K_{unf}[ES]}{[S]}}$$

$$Y_U = \frac{K_M K_{unf}}{K_M + [S] + K_M K_{unf}} = \frac{K_M K_{unf}}{K_M(1 + K_{unf}) + [S]}$$

The loss of activity resulting from protein aggregation or misfolding is described as an irreversible state originating from the unfolded state. The rate of enzyme inactivation is modeled as an exponential decay of the total amount of enzyme, $[E_0]$, which is dependent on time and the fraction of enzyme in the unfolded state.

$$[E]_t = [E]_0 e^{-k_{irrev} \cdot t \cdot Y_U}$$

$$\frac{d[\text{Product}]}{dt} = v_t = \frac{k_{cat}[\text{E}]_t[\text{S}]}{K_M(1 + K_{unf}) + [\text{S}]} = \frac{k_{cat}[\text{E}]_0[\text{S}]}{K_M(1 + K_{unf}) + [\text{S}]} \cdot e^{-k_{irrev} \cdot t \cdot Y_U}$$

Reversible and irreversible folding terms are necessary to fit the experimentally measured kinetics data. Plots of v_i versus substrate concentration were fit to Equation A-1 over a range of ATP concentrations (5, 10, 25, 50, 100, 250, 500 and 1000 μM) at various temperatures (55, 60, 62.5, 65, 67.5 and 70 °C). K_{unf} was incorporated in the fitting scheme as a constant using the values in Table S3. k_{irrev} was treated as being independent of temperature to constrain the values shown in Table 1 for each mutant. Since the impact of k_{irrev} is not appreciably large at temperatures below 70 °C, the fitting scheme cannot determine unambiguously k_{irrev}, and therefore k_{irrev} for AK_{BSUB} Q199R/Q16L was estimated as a value of 0.01 s^{-1} to allow unambiguous fits of K_M and k_{cat}. K_M for ATP at high [AMP] and k_{cat} for ADP production were determined by nonlinear regression fitting to Equation A-1 at each temperature with the respective standard deviations of each value (Table 6.1).

A.5 Independent determination of the rate of irreversible protein loss in solution

The rate of irreversible protein loss can be determined for the enzyme in absence of substrate. The mutant enzymes were diluted to 100 nM and incubated in buffered solution at 50 °C for different periods of time and assayed at room temperature (25 °C) using a continuous coupled assay. A single data point was collected at each time point for each enzyme. The data is best fit to a two phase exponential decay function for AK_{BSUB}Q199R/A193V and AK_{BSUB}Q199R/Q16L. However, a two phase exponential decay shows little improvement over a fit using a single phase exponential decay function for AK_{BSUB}Q199R, AK_{BSUB}Q199R/G213E and AK_{BSUB}Q199R/G214R. There appears to be distinct fast and slow phases. The fast phase occurs over the duration of the enzyme assay (60 seconds), and is approximately equivalent for all enzymes at 50 °C. The slow

phase rate that occurs over a time scale much longer than the assay and varies with the stability of each enzyme, the relative activity for least stable mutant, AK_{BSUB}Q199R, decays the fastest and the most stable, AK_{BSUB}Q199R/Q16L, decays the slowest. It is possible that less surface area in the reaction vessel (a polypropylene, 1.5 mL microcentrifuge tube) is available for the enzyme to anneal to as time progresses. Once the wall of the reaction vessel is coated, the enzyme misfolds at a rate proportional to the fraction of unfolded enzyme. Alternatively, since aggregation is multimolecular the rate dependence should decrease as total concentration decreases.

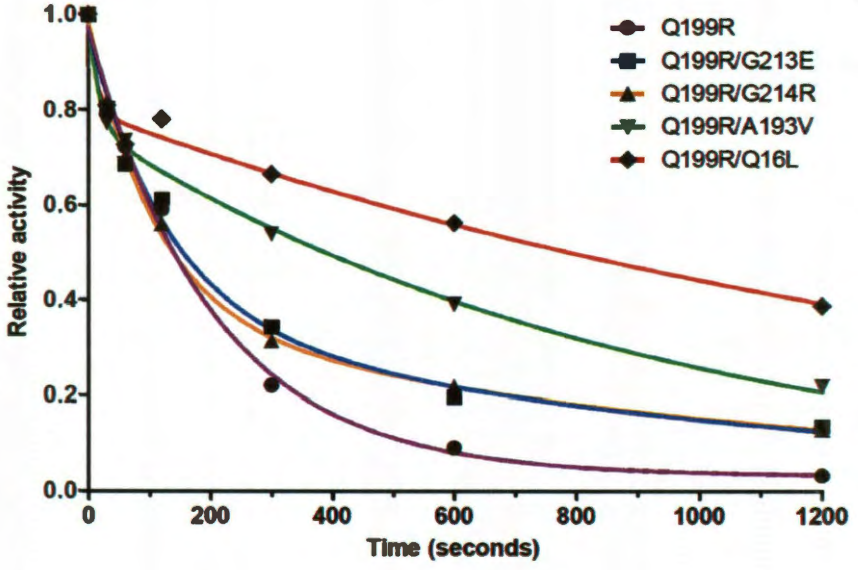


Figure A.1
AK_{BSUB} misfolding at 50 °C is monitored by measuring the residual activity of an enzyme solution as a function of time. There are two clear phases for the most stable enzymes: AK_{BSUB}Q199R/A193V and AK_{BSUB}Q199R/Q16L. This suggests two separate mechanisms of misfolding that can be characterized by exponential rates, one fast and one slow. The slow misfolding rate is highly dependent on stability.

A.6 Thermostability studies using Differential Scanning Calorimetry (DSC)

AK unfolding can be well modeled as an apparent two-state transition although it must be recognized that re-folding is hampered by potential aggregation and partial loss

of an intrinsic zinc ion (Bae & Phillips, 2004; Miller *et al.*, 2010). To assess the extent to which AK and the associated mutants can be reasonably modeled as an apparent two-state transition the thermodynamic parameters for each protein was determined as a function of concentration at 10, 20, and 40 μM and scan rates at 60 and 90 °C/hour (Table S1). In addition, a very low scan rate study of 30 °C/hour was done on mutants AK_{BSUB} Q199R/A193V, AK_{BSUB} Q199R/Q16L and AK_{BSUB} Q199R/G214R. A typical example of the scan rate and concentration dependence of T_m is shown for AK_{BSUB} Q199R/Q16L in Figure A.. The midpoint of the reversible two-state transition T_m, enthalpy change (ΔH), and change in molar heat capacity of the system (ΔC_P) were determined from differential scanning calorimetry (DSC) data analyzed by Origin 7.0 software (MicroCal, LLC, Northampton, USA). Data from experiments at 40 μM AK and 90 °C/hr were used in the population models and were performed in triplicate with errors presented as standard deviations (Table A.1). It should be noted that the thermodynamic parameters calculated using a two-state fit are not used to extrapolate to values at much lower or higher temperatures, but are modeled near to the transition temperature where these values are more accurately determined and are most relevant to our population modeling.

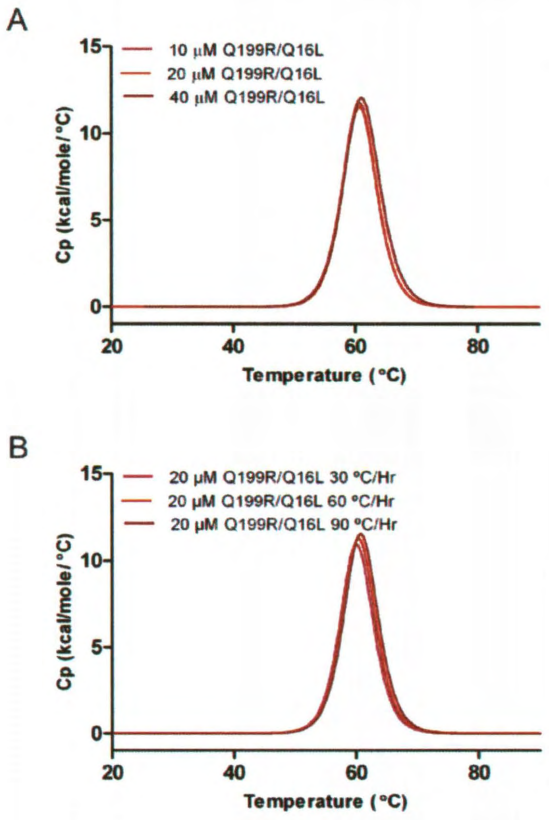


Figure A.2
DSC scans fitted as an apparent two-state transitions show little dependence on (A) concentration of enzyme (10 to 40 μM) and (B) the thermal scanning rate (30-90 °C/hr).

Table A.1

Thermodynamic parameters of AK_{BSUB} mutants obtained from thermal denaturation monitored by DSC.

Enzyme		T_m*, °C	$\Delta H^{\dagger}$, kcal/mol	$\Delta C_p^{\ddagger}$, kcal/mol/deg
B. subtilis WT	40μM 90°/hr scan rate	44.0 ± 0.1	51.9 ± 7.7	0.73 ± 0.2
	10μM 90°/hr scan rate	46.9 ± 0.12	65 ± 1.41	0.83 ± 0.11
	20μM 90°/hr scan rate	47 ± 0.007	60.2 ± 0.8	0.95 ± 0.07
	40μM 90°/hr scan rate	46.8 ± 0.1	66.7 ± 4.3	1.20 ± 0.1
	20μM 60°/hr scan rate	46.39	58.6	0.86
	Standard Deviation	± 0.26	± 3.84	± 0.17
Q199R	10μM 90°/hr scan rate	47.8 ± 0.03	62.7 ± 3.6	0.22 ± 0.1
	20μM 90°/hr scan rate	48.0 ± 0.01	60.2 ± 1.1	0.10 ± 0.01
	40μM 90°/hr scan rate	48.0 ± 0.2	56.6 ± 9.7	0.25 ± 0.1
	20μM 60°/hr scan rate	47.3	56.1	0.3
	Standard Deviation	± 0.33	± 3.33	± 0.09
Q199R/G213E	10μM 90°/hr scan rate	49.5 ± 0.2	68.9 ± 1.3	0.56 ± 0.2
	20μM 90°/hr scan rate	49.7 ± 0.01	67.2 ± 0.1	0.44 ± 0.1
	40μM 90°/hr scan rate	49.5 ± 0.2	67.5 ± 10.2	0.69 ± 0.1
	20μM 60°/hr scan rate	49.0	58.8	0.6
	Standard Deviation	± 0.30	± 4.59	± 0.10
Q199R/T179I	10μM 90°/hr scan rate	50.9 ± 0.01	75.5 ± 3.5	0.64 ± 0.3
	20μM 90°/hr scan rate	51.0 ± 0.01	73.9 ± 5.8	0.66 ± 0.1
	40μM 90°/hr scan rate	51.0 ± 0.3	78.0 ± 4.5	0.64 ± 0.1
	20μM 60°/hr scan rate	50.4	69.2	0.5
	20μM 30°/hr scan rate	49.09	51.33	
	Standard Deviation	± 0.81	± 10.7	± 0.07
Q199R/A193V	10μM 90°/hr scan rate	51.9 ± 0.04	82.8 ± 3.0	1.01 ± 0.07
	20μM 90°/hr scan rate	52.0 ± 0.1	82.9 ± 0.6	1.09 ± 0.04
	40μM 90°/hr scan rate	52.8 ± 0.3	81.0 ± 9.6	1.09 ± 0.1
	20μM 60°/hr scan rate	51.7	82.5	1.1
	20μM 30°/hr scan rate	51.3	103.0	
	Standard Deviation	± 0.55	± 9.29	± 0.04
Q199R/Q16L	10μM 90°/hr scan rate	60.8 ± 0.02	88.1 ± 0.7	1.1 ± 0.1
	20μM 90°/hr scan rate	60.8 ± 0.04	86.1 ± 0.1	1.1 ± 0.05
	40μM 90°/hr scan rate	60.9 ± 0.2	79.2 ± 15.0	1.09 ± 0.2
	20μM 60°/hr scan rate	60.5	81.9	1.1
	20μM 30°/hr scan rate	60.2	79.1	
	Standard Deviation	± 0.29	± 4.07	± 0.01

*T_m is the thermal midpoint of a transition
†ΔH is the calorimetric heat change
‡ΔC_p is the heat capacity change

The change in Gibbs free energy (ΔG_{unf}) is defined as the difference between free energy in the unfolded (G_u) and folded (G_f) state. The calculated values of ΔG_{unf} in Table II and S2 were determined at various temperatures using the thermodynamic parameters in Table S1 and the following modified Gibbs-Helmholtz equation (Becktel & Schellman, 1987):

$$\Delta G_{unf}(T) = \Delta H(T_m)\left(1-\frac{T}{T_m}\right) - \Delta C_p\left(T\ln\frac{T}{T_m} + (T_m - T)\right) \qquad \text{A-3 (4-11)}$$

where $\Delta H(T_m)$ is the enthalpy change at the transition temperature. The following equation describes the relation between unfolding equilibrium constants (K_{unf}), ΔG_{unf}, folding rates, and the ratio of unfolded and folded enzyme:

$$K_{unf} = e^{-\Delta G_{unf}/RT} = \frac{k_u}{k_f} = \frac{[U]}{[E]} \qquad \text{A-4}$$

Table A.2

Calculated Gibbs free energy change, ΔG_{unf} (kcal/mol), at respective temperatures for each mutant with standard deviations.*†

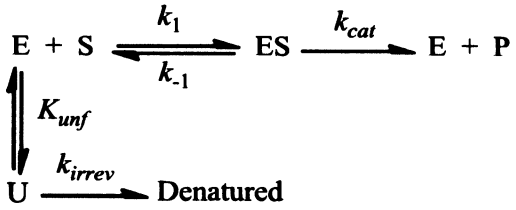
Enzyme		Temperature					
		55	60	62	65	68	70
Q199R	10μM	-1.74 ± 0.07	-2.89 ± 0.12	-3.36 ± 0.13	-4.10 ± 0.17	-4.85 ± 0.20	-5.37 ± 0.22
	20μM	-1.60 ± 0.01	-2.70 ± 0.02	-3.15 ± 0.02	-3.86 ± 0.01	-4.60 ± 0.01	-5.10 ± 0.01
	40μM	-1.83 ± 0.13	-3.07 ± 0.22	-3.59 ± 0.26	-4.40 ± 0.31	-5.24 ± 0.38	-5.82 ± 0.42
	20μM 60 °/hr	-1.68	-2.74	-3.19	-3.87	-4.58	-5.06
	s.d.	±0.10	±0.17	±0.20	±0.25	±0.31	±0.35
Q199R/G213E	10μM	-1.46 ± 0.04	-2.47 ± 0.10	-2.88 ± 0.14	-3.50 ± 0.18	-4.12 ± 0.23	-4.54 ± 0.26
	20μM	-1.33 ± 0.03	-2.28 ± 0.04	-2.67 ± 0.05	-3.24 ± 0.06	-3.82 ± 0.07	-4.21 ± 0.07
	40μM	-1.25 ± 0.02	-2.17 ± 0.36	-2.54 ± 0.43	-3.21 ± 0.65	-3.68 ± 0.64	-4.07 ± 0.71
	20μM 60 °/hr	-1.37	-2.28	-2.65	-3.21	-3.78	-4.17
	s.d	±0.09	±0.12	±0.14	±0.14	±0.19	±0.20
Q199R/T179I	10μM	-1.19 ± 0.03	-2.33 ± 0.04	-2.79 ± 0.04	-3.50 ± 0.06	-4.23 ± 0.08	-4.72 ± 0.09
	20μM	-1.12 ± 0.01	-2.22 ± 0.01	-2.60 ± 0.01	-2.89 ± 0.02	-3.80 ± 0.04	-4.03 ± 0.04
	40μM	-1.17 ± 0.01	-2.30 ± 0.30	-2.77 ± 0.37	-3.49 ± 0.47	-4.22 ± 0.58	-4.72 ± 0.66
	20μM 60 °/hr	-1.13	-2.12	-2.52	-3.14	-3.78	-4.21
	s.d.	±0.03	±0.09	±0.13	±0.30	±0.25	±0.35
Q199R/G214R	10μM	-0.96 ± 0.50	-2.19 ± 0.13	-2.70 ± 0.18	-3.47 ± 0.23	-4.26 ± 0.30	-4.80 ± 0.35
	20μM	-0.93 ± 0.08	-2.14 ± 0.18	-2.63 ± 0.22	-3.40 ± 0.29	-4.17 ± 0.36	-4.70 ± 0.40
	40μM	-0.99 ± 0.03	-2.25 ± 0.08	-2.77 ± 0.11	-3.58 ± 0.18	-4.38 ± 0.21	-4.93 ± 0.24
	20μM 60 °/hr	-1.00	-2.12	-2.58	-3.28	-3.99	-4.47
	s.d.	±0.03	±0.06	±0.08	±0.13	±0.16	±0.19
Q199R/A193V	10μM	-0.80 ± 0.02	-2.15 ± 0.07	-2.72 ± 0.09	-3.59 ± 0.13	-4.49 ± 0.16	-5.10 ± 0.19
	20μM	-0.79 ± 0.02	-2.16 ± 0.01	-2.73 ± 0.01	-3.61 ± 0.02	-4.51 ± 0.02	-5.14 ± 0.02
	40μM	-0.56 ± 0.08	-1.88 ± 0.20	-2.43 ± 0.26	-3.28 ± 0.34	-4.17 ± 0.43	-4.77 ± 0.49
	20μM 60 °/hr	-0.90	-2.22	-2.79	-3.67	-4.57	-5.19
	s.d.	±0.14	±0.15	±0.16	±0.18	±0.18	±2.45
Q199R/Q16L	10μM	1.48 ± 0.01	0.21 ± 0.01	-0.32 ± 0.01	-1.14 ± 0.01	-1.98 ± 0.02	-2.56 ± 0.03
	20μM	1.45 ± 0.01	0.22 ± 0.01	-0.30 ± 0.01	-1.1 ± 0.01	-1.93 ± 0.01	-2.50 ± 0.02
	40μM	1.35 ± 0.30	0.22 ± 0.09	-0.25 ± 0.01	-0.99 ± 0.14	-1.75 ± 0.28	-2.28 ± 0.37
	20μM 60 °/hr	-1.32	-0.14	-0.35	-1.12	-1.91	-2.45
	s.d.	±0.08	±0.04	±0.04	±0.07	±0.10	±0.12

*ΔG_{unf} at the designated temperatures for each mutant is calculated using Equation S4 given the thermal midpoint (T_m), enthalpy change at the midpoint ($\Delta H(T_m)$), and the change in heat capacity (ΔC_p) determined by DSC and listed in Table S1.
†The standard deviation is determined for three calculated values of ΔG_{unf}.

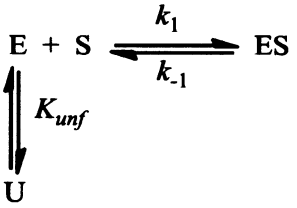
A.7 Stabilization in the presence of ligand

Enzyme kinetics data fitted to Equation 1 in the main text accounted for the fraction of unfolded protein (Y_U) at a given concentration of free substrate and temperature. This function (Equation A-5, Equation 6-7 in the main text) can be used to

estimate the amount of ligand required to increase the melting temperature and hence the fraction of enzyme folded at a given temperature. A similar function was first derived by Schellman and utilized to predict protein stabilization by Pace and McGrath (Schellman, 1975; Pace & McGrath, 1980).



Scheme I



Scheme II

The equation describing the effect of ligand binding on the fraction of unfolded enzyme (Y_U) is general to either Scheme I or II and can thus be applied to a range of methods such as enzyme kinetics (I) and two-state unfolding monitored by DSC in the presence of ligand (II). The treatment of the Michaelis-Menten constant (K_M) in the derivation of Y_U makes it interchangeable with the dissociation constant (K_D=K_L^{-1}); however, it should be noted that the rates that define the equilibrium constants differentiate these two distinct concepts.

$$K_M = K_L^{-1} = \frac{[E][S]}{[ES]} \text{ whereas } K_M = \frac{k_{-1} + k_{cat}}{k_1} \approx K_L^{-1} = \frac{k_{-1}}{k_1}$$

The relation above can be approximated only if the catalytic rate (k_{cat}) is much slower than the dissociation (k_{-1}) rate. Therefore, the expression for the dependence of the fraction of unfolded enzyme on the ligand concentration ([S]), unfolding equilibrium

constant for the apoprotein at a given temperature (K_{unf}(T)), and the apparent affinity of the enzyme for ligand (K_M or K_L) can be written as:

$$Y_U(\mathrm{T}) = \frac{K_M K_{unf}(\mathrm{T})}{K_M(1+K_{unf}(\mathrm{T}))+[S]} = \frac{K_{unf}(\mathrm{T})}{1+K_{unf}(\mathrm{T})+K_L[S]} \qquad \text{A-5}$$

The transition temperature in the presence of ligand (T_m) is greater than in its absence (T_0) due a shift in a fraction of total enzyme away from the unfolded state toward the ligand-bound state. We can quantitatively validate this function by comparing the T_m determined at a specific concentration of ligand with that calculated with the equation above using only the ligand binding constant (K_L) determined with Equation A-6 and the Gibbs free energy for apoprotein unfolding as a function of temperature (Table A.2).

Ligand affinity was measured using the equation below for the estimation of K_{ligand} or K_L for tight binding from Brandts and Lin (1990).

$$K_L(\mathrm{T_m}) = \frac{\exp\left\{\frac{-\Delta H(\mathrm{T_0})}{R}\left(\frac{1}{\mathrm{T_m}}-\frac{1}{\mathrm{T_0}}\right)+\frac{\Delta C_p}{nR}\left(\ln\frac{\mathrm{T_m}}{\mathrm{T_0}}+\frac{\mathrm{T_0}}{\mathrm{T_m}}-1\right)-1\right\}}{[L]_{T_m}} \qquad \text{A-6}$$

where, $[L]_{T_m}$ is the concentration of free ligand at T_m, L_t and P_t are total ligand and protein concentrations, respectively, in solution, and

$$[L]_{T_m} = L_t/2 \text{ if } L_t \le P_t \text{ and } [L]_{T_m} = L_t - P_t/2 \text{ if } L_t \ge P_t$$

The fraction of unfolded protein at the transition temperature is 0.5.

$$Y_U(\mathrm{T_m}) = 0.5 = \frac{K_{unf}(\mathrm{T_m})}{1+K_{unf}(\mathrm{T_m})+K_L[S]}$$

$$1+K_{unf}(T_m)+K_L[S]=2K_{unf}(T_m)\Rightarrow 1+K_L[S]=K_{unf}(T_m)=e^{-\Delta G_{unf}(T_m)/RT_m}$$

$$\ln(1+K_L[S])=-\Delta G_{unf}(T_m)/RT_m$$

The Gibbs free energy for unfolding is treated as linear function of temperature within this regime (not shown, data in Table A.2).

$$\Delta G_{unf}(T)=\text{slope}_{\Delta G_{unf}}T+\text{intercept}_{\Delta G_{unf}}$$

$$T_m=\frac{\text{intercept}_{\Delta G_{unf}}}{-R\ln(1+K_L[S])-\text{slope}_{\Delta G_{unf}}} \qquad \text{A-7}$$

By satisfying this function, the transition temperature can be determined in the presence of a given concentration of substrate with a predetermined binding constant. The predicted transition temperatures are very close to observed values (see Table 6.3 and Table A.3).

Table A.3

Comparison between transition temperatures of AK_{BSUB} Q199R/A193V monitored by DSC in the presence of various ligands and those predicted by Equation A-7 (T_0=52.0 °C for apoprotein).

Ligand	K_L*, M	[S]†, M	T_m‡, °C	
			Observed	Predicted
ATPγS	7.9×10^{2}	1×10^{-3}	53.4	54.6

*K_L is the binding constant determined at 20 μM protein at a scan rate of 90 °C/hr.
†Estimation of the concentration of free ligand in the presence of 20 μM protein.
‡T_m is the thermal midpoint of a transition where approximately 50 % of the ensemble is folded and bound in the presence of ligand.

It is important to note that AK is active and will carry out product formation during DSC in the presence of ATP and AMP or ADP. To demonstrate stabilization, the transition state analog Ap_5A was used. As outlined by Brandts and Lin (1990) under very tight binding conditions with lower than stoichiometric concentrations of substrate, the

transition temperature of the population will shift and distribute into two populations with two melting temperatures: a T_0 for the apoprotein (unbound) and a T_m that corresponds to the enzyme-ligand complex melting temperature (Brandts & Lin, 1990). Figure 6.4 of the main text shows two distinct peaks: unbound AK_{BSUB} Q199R/A193V and bound to the transition state analog Ap_5A.

We performed DSC measurements at 0.25:1, 0.5:1 and 1:1 stoichiometric ratios of Ap_5A to AK_{BSUB} Q199R/A193V (Figure 6.4). Ap_5A increased AK_{BSUB} Q199R/A193V stability by 12.3 °C. AK_{BSUB} Q199R/A193V showed little change in T_m as a function of Ap_5A concentration for the bound state as would be expected (Brandts & Lin, 1990). Knowing the affinity of Ap_5A for AK and the T_m (Table A.3) allowed us to compare the measured stabilization to that predicted from Equation A-7 and is in excellent agreement with the observed data (Table A.3). Conversely, this data can be used predict the amount of Ap_5A that is required to shift the T_m of AK_{BSUB} Q199R/A193V to 63.4 °C. Again, using Equation A-7, the predicted concentration of Ap_5A would be 12 μM in excellent agreement with the experimentally measured data.

In addition to measuring the extent of Ap_5A mediated stabilization of AK_{BSUB} Q1993R/A193V, the weaker binding case of ATP binding to the enzyme was also estimated, although the activity of the enzyme limits interpretation of the data to an approximation. In the weaker binding case of ATP, ATPγS and AMP, there is little appreciable shift in T_m until excess ligand is added. Unfortunately, the "non-hydrolyzable" analog ATPγS does not inhibit AK activity (not shown) meaning that reaction containing either ATP or ATPγS and AMP will continue to equilibrate during

the experiment. In addition, even the expected 1-2% contamination of ATP with ADP would mean that a concentration of 10-20 μ 1 ADP is present in the cell containing 20 μM AK and hinders a clear interpretation of DSC data since the enzyme will be carrying out catalysis throughout the experiment.

As more ligand is added, the transition temperature begins to increase. As shown by Brandts and Lin (1990) a single transition protein with two binding sites for a small molecule (AMP and ATP in this case) with comparable affinities will distribute into two peaks (T_0 = unbound) and a new peak with higher stability that is the doubly bound state (Brandts & Lin, 1990). The singly bound state is not observed at the statistical fraction that one expects as there is strong cooperativity to populate the more stable doubly liganded population as temperature increases (Brandts & Lin, 1990). As shown in Figure 6.4 and Table A.4, this effect is also what is observed for AK_{BSUB} Q199R/A193V in the presence of excess substrate. Based on the simulations done by Brandts and Lin, the shift and separations in T_0 and T_m are appropriate for the expected estimated affinities of the ligands AMP and ATP. If we make the assumption that the T_m measured above is the temperature at which 1/2 the protein is folded and potentially bound at a particular substrate concentration, then we can compare this to values calculated using Equation A-7 and K_L and ΔG_{unf} as a function of temperature determined by DSC. Again, as shown in Table 6.3 and Table A.3, there is excellent agreement between the predicted and measured stability.

An important series of caveats exist here: (1) this analysis assumes a largely two-state model or that the data can be well modeled by a two-state approximation, (2) ligand

will only bind to folded AK, (3) all ΔC_p values are temperature independent, (4) activity coefficients are unity (Brandts & Lin, 1990).

Table A.4

Shifts in transition temperatures for AK_{BSUB} Q199R/A193V in the presence of various ligands at different concentrations as monitored by DSC.

Ligand	T_m, °C	Change relative to T_0
No ligand	52.0 ± 0.1	-
5 μM Ap_5A *	64.2 ± 0.1	+12.2
10 μM Ap_5A*	64.2 ± 0.1	+12.2
20 μM Ap_5A *	64.3 ±0.03	+12.3
1 mM AMP	53.4	+1.4
1 mM ATP	59.1	+7.1
1 mM ATPγS	53.4	+1.5
1 mM AMP + 0.25 mM ATPγS	54.9	+3.0
1 mM AMP + 1.0 mM ATPγS	57.0	+5.0

*Ap_5A = transition state analog for AK, P^1, P^5-di(adenosine-5') pentaphosphate

A.8 Population Dynamics

A numerical solution with discrete time steps was used to simulate the population dynamics of an exponentially growing polymorphic population of bacteria. A geometric growth model was assumed with given initial conditions and temperature dependent reproductive rates for each respective mutant.

$$P_i(t+1) = P_i(t) + r_i(t)P_i(t) \qquad \text{A-8}$$

P_i is the frequency of a population i growing at a rate of r_i at generation t. The growth rate (r_i) can be a value between 0 and 1. If the $r_i = 0$, then the population size at the next generation (t+1) is the same as the previous (t). However, if $0 < r_i \leq 1$, then there is a fractional increase in the population size or a full doubling of the population over the generation in the case where $r_i = 1$. For a polymorphic population, each subpopulation has a function with its respective growth rates. Each of the six allelic subpopulations enriched by natural selection and isolated from the experimental culture were modeled in this way.

The simulation is limited to predicting evolutionary outcomes of enzymes investigated by *in vitro* methods and cannot predict appearance of novel mutations. Each mutation alters the biochemical properties of the enzyme leading to downstream effects that alter the average growth rate, r_i, of the organism (fitness) within the subpopulation. The general numerical solution for a mutant population takes Equation A-8 and includes rates of mutations forming the allele of interest (P_i) and mutations that form new alleles (P_j) from the starting P_i allele:

$$P_i(t+1) = P_i(t)(1+r_i(t)) + \sum_j m_{ji} r_j(t) P_j(t) - \sum_j m_{ij} r_i(t) P_i(t) \qquad \text{A-9}$$

where m_{ji} is the rate of substitution which creates allele i from j, and m_{ij} is the reciprocal mutation creating allele j from i. The mutation rate per base per replication used in the model was estimated at 5E-10 based on rates measured for mutations in *E. coli* (Drake *et al.*, 1998). In addition to mutation rate, estimates were made for the initial conditions that include the population size for the first generation ($t = 0$) and the limiting population size determined by Couñago *et al.* (2006).

A turbidostat regulates the flux of media to maintain a maximum cell density (0.4-0.6 OD) and constant volume (0.5 L) to attain a steady-state population. Excess media, containing replicating and quiescent cells, is removed as waste. The population cap is modeled by the redefining the sizes of each subpopulation when the total population size (P_{total}) surpasses the maximum value (P_{max}).

$$P_{total} = \sum_k P_k(t) = P_{Q19R} + P_{Q19R/Q16L} + P_{Q19R/T179} + P_{Q19R/A198} + P_{Q19R/G21E} + P_{Q19R/G21R} \qquad \text{A-10}$$

The population is assumed to consist of the six isolated strains identified by random sampling, all of which reached levels of 5 % or greater within the population

(Couñago *et al.*, 2006). If the total population is larger than the allowed value, $P_{total} > P_{max}$, then $P_i(t+1) = P_i(t+1)(P_{max} / P_{total})$. The number of organisms within each subpopulation is multiplied by the ratio of the maximum size to the sum of all subpopulations, redefining the subpopulation size for the next time step. The numerical simulation is iterated over 1500 generations (t = 1500), which roughly translates to a generation occurring every 30 minutes over the course of the experiment (1 month) where the temperature was increased from 55 to 70 °C. When the growth rate is less than 1 (r_i < 1), the fraction of the population replicating at a given time step is equivalent to the growth rate at that time step. Activity profiles for each enzyme were determined, and multiple fitness functions incorporating kinetic parameters were evaluated on the basis of reproducing the known evolutionary outcomes in the context of the population simulation.

A.9 Numerical simulation: varying substrate concentration (250, 500, 1000, 2000 μM)

Substrate stabilization of AK_{BSUB} contributes to the population dynamics observed in the numerical solution and presumably during natural selection. As seen in Figure A.3, increases in the concentration of ATP leads to prolonged persistence of Q199R (purple), shifting the transition between allelic dominance to later generations. The increase in substrate also increases the activity of all enzymes, eventually making them equally competitive as their activity surpasses the threshold. The unfolding midpoints of almost all mutations (with the exception of Q199R/Q16L) are below the beginning experimental temperature. Q199R has a T_m of 47 °C, but the mutant strain was isolated in the culture up to 62 °C. Substrate stabilization in the model may be accounting

for various alternative mechanisms of stabilization such as macromolecular crowding and chaperones. *In vivo*, a combination of these factors will produce additional stability, but in the assay an increase in [S] effectively produces a comparable overall extent of stabilization and retention of AK activity.

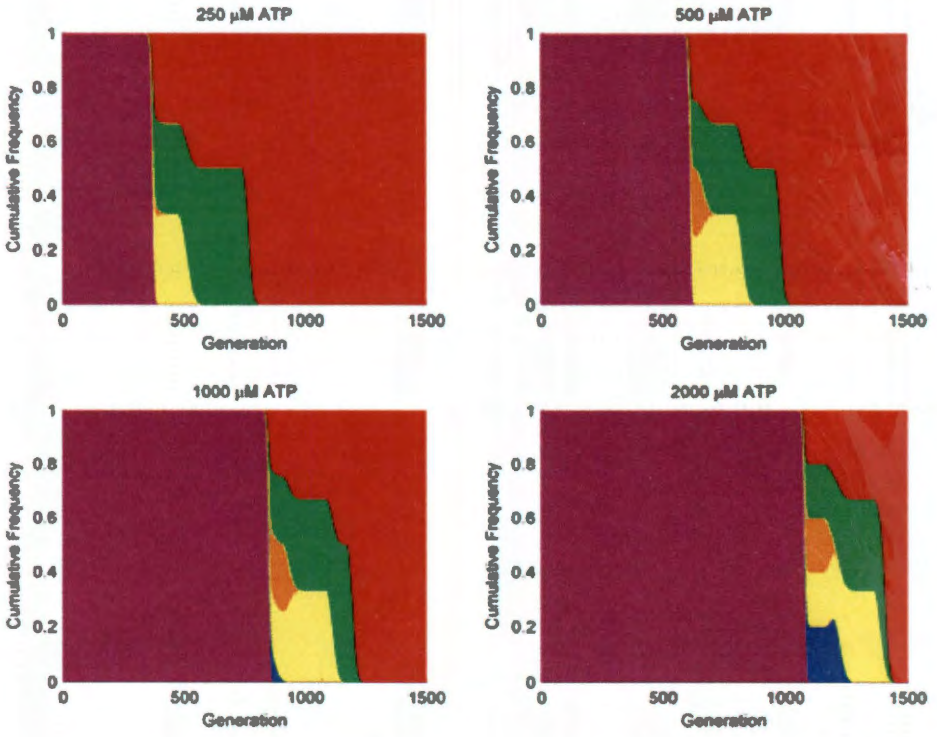


Figure A.3
Changes in substrate concentration (ATP) illustrate how the evolutionary fates of the microbial population are directly correlated to the folded fraction of AK. Numerical solutions were performed with various substrate concentrations, 250, 500, 1000, and 2000 μM ATP. *In vivo*, the fraction folded will also depend on the activity of chaperones and osmolytes that would stabilize the native protein. At higher concentrations of ATP, the folded fraction of AK_{BSUB} Q199R increases delaying the rise of the double mutants to much later in the experiment. *In vivo*, the AK_{BSUB} Q199R population begins declining at ~500 generations (~62 °C) and is largely absent by ~900 generations (~64 °C). The *in vivo* data are comparable to the data shown for 500-1000 μM ATP in good agreement with estimated concentrations of ATP *in vivo*. The colors correspond to the color scheme used to identify the positions of the mutations in Figure 1. 0, 500, 1000 and 1500 generations correspond to 55, 60, 65 and 70 °C, respectively.

A.10 Numerical simulation: various activity thresholds (no threshold, 10, 8.75, and 7 μM ADP/min/nM)

The observed transient subpopulations (Q199R/T179I, Q199R/G213E, and Q199R/G214R) cannot be observed in the absence of an activity threshold, even when implementing a hyperbolic relation between activity and fitness (Kacser & Burns, 1981; Dykhuizen *et al.*, 1987; Lunzer *et al.*, 2005). AK_{BSUB} Q199R/Q16L has much greater activity, and when activity is used as a direct proxy for fitness, as can be seen in the first panel of Figure A.4, the greater activity of mutant Q199R/Q16L leads to an early and direct transition from Q199R. The lower the threshold drops (7 compared to 8.75 μM ADP/min/nM), the more other mutants appear and persist in subsequent generations before their activities falls below the lowered threshold. At a relatively high threshold (13 μM ADP/min/nM) only two other mutants arise to prominence after the fall off of AK_{BSUB} Q199R and before the dominance of AK_{BSUB} Q199R/A16L (Figure A.4). A threshold of 8.75 μM ADP/min/nM was used in the final model and provides the best simulation of the observed data shown in Figure 6.3D of the main text. The double mutant to first fall below the threshold (AK_{BSUB} Q199R/G213E) is the most sensitive to changes in the threshold. The fractional representations of Q199R/G213E in the population for thresholds of 8, 8.75, and 9 μM ADP/min/nM are 0.200, 0.111, and 0.056, respectively. By comparison, the fractions of the population made up of Q199R/G214R at these thresholds are 0.236, 0.248, and 0.249, respectively. The activity of Q199R falls below the threshold of 8.75 μM ADP/min/nM at generation 722 in the simulation. It is not until generation 840 that Q199R falls below a cumulative 95% of the population. The precipitous decrease in the fraction of Q199R single mutant after generation 840, which corresponds to a temperature of 63.4 °C in the turbidostat, correlates well with the

midpoint of the decrease in frequency determined through sequencing from the continuous population at 62 °C (Couñago *et al.*, 2006).

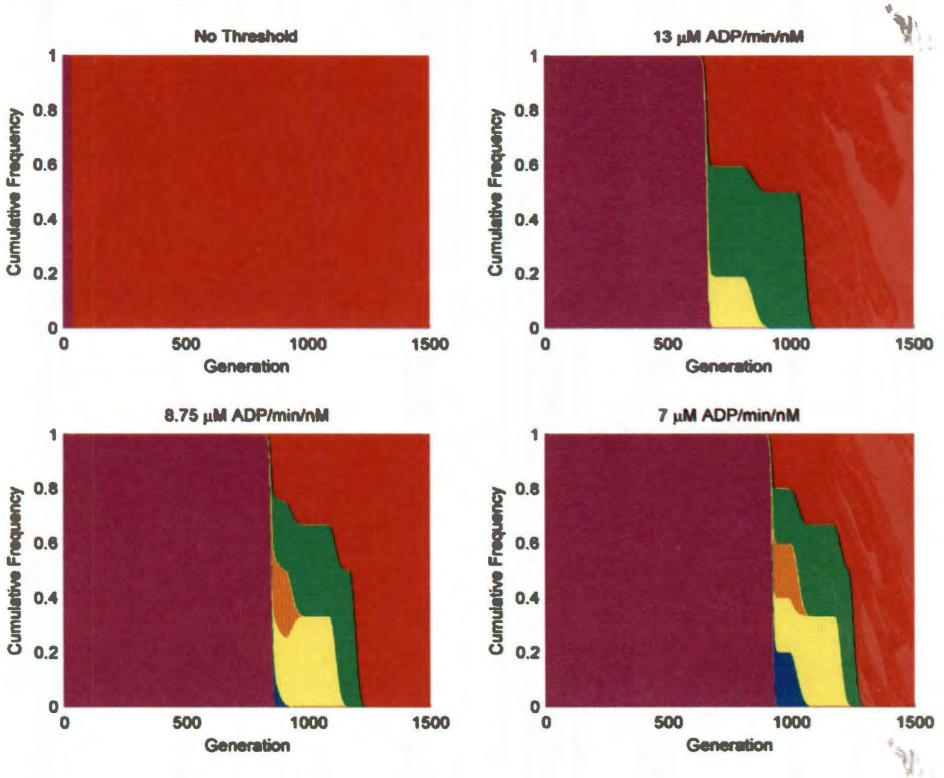


Figure A.4
Evolutionary outcomes in the numerical simulation are dependent on the threshold chosen. Higher thresholds restrict the appearance and persistence of mutants with lower activities while accelerating the demise of the less fit progenitor. Decreasing the threshold permits mutants with lower activities to remain competitive for longer. 0, 500, 1000 and 1500 generations correspond to 55, 60, 65 and 70 °C, respectively.

A.11 Numerical simulation: with and without differences in substitution rate forming AK_{BSUB}Q199R/Q16L (5E-10 versus 5E-11)

The fitness function forces a stable polymorphism by constraining fitness to be constant and equivalent when activities are above the threshold. Application of the activity threshold or a hyperbolic fitness function creates situations where fitnesses are equivalent or very similar. Under these conditions, disparate mutation rates can affect the

frequencies of equivalently fit alleles within a polymorphic population. Within the numerical solution, the ratio of mutation rates between mutants of the same fitness corresponds to the relative frequency between each mutant. A tenfold decrease in substitution rate for Q16L (transversion) relative to the substitutions needed to generate the other point mutants (transitions) translates to a decrease in frequency in of the Q199R/Q16L mutant, and also an increase in frequency of the viable mutants due to the decrease in clonal interference attributed to AK_{BSUB} Q199R/Q16L (Figure A.5).

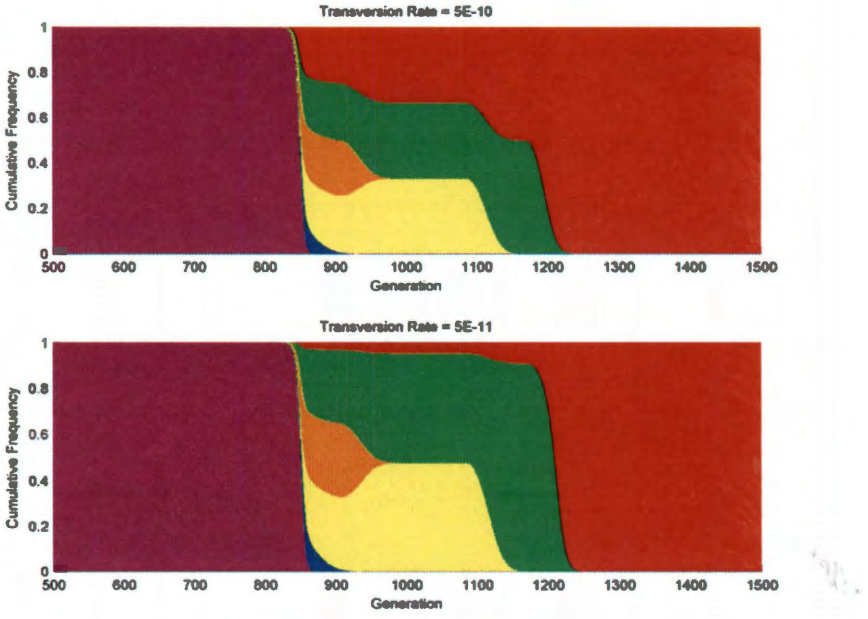


Figure A.5
Mutation rate contributes to allelic frequency under circumstances where competing strains have similar or effectively identical fitnesses. The Q16L mutation is caused by a transversion, which typically occurs at a lower rate than transitions. If the transition bias is accounted for, AK_{BSUB} Q199R/Q16L remains at a lower frequency than in the absence of a bias. Generations 500, 1000 and 1500 correspond to 60, 65 and 70 °C, respectively.